1-1-1993

Essays on the Economics of Education

Emily P. Hoffman, Editor
Western Michigan University

Follow this and additional works at: https://research.upjohn.org/up_press

Part of the [Education Economics Commons](#)

Citation

Hoffman, Emily P., ed. 1993. Essays on the Economics of Education. Kalamazoo, MI: W.E. Upjohn Institute for Employment Research. <https://doi.org/10.17848/9780880995542>

This work is licensed under a [Creative Commons Attribution-Noncommercial-Share Alike 4.0 License](#).

This title is brought to you by the Upjohn Institute. For more information, please contact repository@upjohn.org.

ESSAYS on the ECONOMICS of EDUCATION

Emily P. Hoffman, *Editor*

ESSAYS on the ECONOMICS of EDUCATION

Emily P. Hoffman, *Editor*

1993

W. E. UPJOHN INSTITUTE for Employment Research
Kalamazoo, Michigan

Library of Congress Cataloging-in-Publication Data

Essays on the economics of education / Emily P. Hoffman, editor.

p. cm.

"Papers . . . originally presented during the 1991-1992 academic year as the 28th annual Public Lecture Series organized by the Department of Economics at Western Michigan University"—CIP galley.

Includes bibliographical references and index.

ISBN 0-88099-142-9. — ISBN 0-88099-141-0 (pbk.)

1. Education—Economic aspects—United States—Congresses.
2. Education—Finance—United States—Congresses.
3. Socially handicapped children—Education—United States—Finance—Congresses.
4. Educational equalization—Congresses. I. Hoffman, Emily P.

LC66.E87 1993

379.1'1' 0973—dc20

93-47944

CIP

Copyright © 1993

W. E. Upjohn Institute for Employment Research

300 S. Westnedge Avenue

Kalamazoo, Michigan 49007-4686

The facts presented in this study and the observations and viewpoints expressed are the sole responsibility of the authors. They do not necessarily represent positions of the W. E. Upjohn Institute for Employment Research.

Cover design by J. R. Underhill

Index prepared by Shirley Kessel.

Printed in the United States of America.

Acknowledgments

The papers in this collection, "Essays on the Economics of Education" were originally presented during the 1991-1992 academic year as the 28th annual Public Lecture Series organized by the Department of Economics at Western Michigan University. The lecture series was supported by the W. E. Upjohn Institute for Employment Research and the College of Arts and Sciences of Western Michigan University. Kevin Hollenbeck and Robert G. Spiegelman of the Upjohn Institute and Werner Sichel, Raymond Zelder, and William Bosshardt, my colleagues in the Economics Department, provided much appreciated encouragement, help, and support. I thank Judith K. Gentry of the Upjohn Institute for managing the production of this book. Last, but not most, I am grateful to the authors of these papers for making this volume possible.

Emily P. Hoffman
Kalamazoo, Michigan
May 1993

The Authors

Nabeel Alsalam is a researcher at the National Center for Educational Statistics.

Mary Jean Bowman is Emeritus Professor of Economics and Education at the University of Chicago.

W. Lee Hansen is Professor of Economics and Educational Policy Studies at the University of Wisconsin-Madison.

Eric A. Hanushek is Professor of Economics and Political Science at the University of Rochester.

Emily P. Hoffman is Professor of Economics at Western Michigan University.

Estelle James is Professor of Economics at the State University of New York at Stony Brook and Senior Economist at the World Bank

Henry M. Levin is David Jacks Professor of Education and Economics and Director of the Center for Educational Research at Stanford University.

Robert H. Meyer is Assistant Professor at the Harris Graduate School of Public Policy Studies at the University of Chicago.

CONTENTS

Introduction	1
Emily P. Hoffman <i>Western Michigan University</i>	
The Economics of Education for At-Risk Students	11
Henry M. Levin <i>Stanford University</i>	
Can Equity Be Separated from Efficiency in School Finance Debates?	35
Eric A. Hanushek <i>University of Rochester</i>	
Can Schools Be Held Accountable for Good Performance?	75
Robert H. Meyer <i>University of Chicago</i>	
College Choice, Academic Achievement and Future Earnings	111
Estelle James <i>State University of New York and World Bank</i> Nabeel Alsalam <i>National Center for Educational Statistics</i>	
The Financial Squeeze on Higher Education Institutions and Students	139
W. Lee Hansen <i>University of Wisconsin</i>	
The Economics of Education in a World Of Change	163
Mary Jean Bowman <i>University of Chicago Emeritus</i>	
References	175
Index	177

Introduction

Emily P. Hoffman
Western Michigan University

As the twentieth century comes to a close, educational quality and financing have emerged as among the most important social and economic issues of concern to this nation. The Economics of Education refers to the study of how resources are allocated to achieve educational goals. Questions of equity, efficiency, and effectiveness have been addressed to all levels of formal education—from early childhood education to higher education. This series of papers by prominent economists who have specialized in educational issues brings an economic perspective to several major questions being asked of the educational system in this country and contributes to the national debate.

Three of the papers address elementary and secondary education. The thrust of the Henry M. Levin paper is to examine the educational opportunities afforded “at risk” children at the primary level and to describe an alternative approach being tested in a number of locations across the country. Eric A. Hanushek takes on the issue of whether or not equity in public support of elementary and secondary schooling will improve student outcomes. Robert H. Meyer points out that performance-based incentives for elementary and secondary schools must be carefully considered and implemented.

Two of the papers focus on higher education issues. Estelle James and Nabeel Alsalam examine the economic payoffs to various higher education choices. They address the question of whether an institution's reputation “pays off” in the labor market for individuals who graduate from it. W. Lee Hansen documents the cyclical nature of the mission of higher education as it has swung from emphases on equity of access to quality and back again.

In the final paper, Mary Jean Bowman philosophizes about the dynamics of educational policy and practice. Curriculums and instructional approaches that are state of the art today may be obsolete tomorrow. Therefore, as policymakers grapple with quality and financing

issues, they need to identify and heed the fundamental and unchanging purposes of education.

In "Economics of Education for At-Risk Students," Levin expresses his concern over the high student dropout rates from our current school system. As a remedy, he offers and explains his accelerated schooling concept. Levin defines at-risk students as those who are unlikely to succeed in existing schools—over half of whom do not graduate—because they do not have the background on which success in that school system is based.

Levin feels that schools have taken the wrong approach in their attempts to help the weak student; he charges that the prevalent remedial or compensatory education with repetitive drill *slows* learning of at-risk students. He feels that the current system of remedial education does this by lowering the expectations of the teachers, which lowers the achievements of the students. Indeed, both the teachers and the students are stigmatized by the remedial labelling. Levin believes that failing students need to be challenged and accelerated, *not* given remedial work. Levin finds that the methods that work well with gifted students also work well with at-risk students—high expectations and stimulating material results in success for at-risk students.

Levin recommends three major changes in the American school system to allow the development of accelerated education. First, the schools must have a clear objective of bringing at-risk children into the mainstream of education, not allowing them to languish in the backwater. Second, there must be school-site empowerment. This means that the decisions affecting the operation of the school must not be passed down from a remote central administration, but must be made by those directly involved in the educational process—the school administrators, the teachers, the parents, even the students. Of course, there must be some overall system of accountability, with rewards commensurate with performance. Third, the schools must build on the strengths of the teachers, the parents, and the students—not dwell on their weaknesses as an excuse for failure.

Levin describes the Stanford Accelerated Schools Project, where the methods that he proposes have been tried. He reviews the pattern of success of these "accelerated schools," which rely on an enrichment strategy rather than the "dumbed-down" rote repetition of conventional

remedial education. He claims that these methods result in a high rate of learning by at-risk students, citing examples of impressive gains in achievement test levels among students from the most deprived socioeconomic background in schools that have adopted the methods he has developed. Amazingly, this progress was achieved by reorganizing the schools, not by increasing the expenditures per pupil.

Levin presents estimates of the proportion of students who are at-risk, and the implications of this for the future quality of our labor force. Besides the general societal benefits, Levin claims that the monetary benefits, in terms of less need for social services, combined with higher incomes and the resulting higher tax revenues, more than compensate for the costs of the accelerated school programs. He summarizes cost-benefit studies of investments in at-risk students and finds a range of \$3 to \$6 in benefits for each \$1 in cost.

If the changes Levin recommends can be widely adopted and the pattern of success can be repeated, all of us—students, parents, teachers, taxpayers, employers, every member of society—will be much better off.

In “Can Equity Be Separated from Efficiency in School Finance Debates,” Eric Hanushek contends that just spending more money on the existing school system does not guarantee improved student achievement. He tries to apply the economic concept of “production function,” which might be more familiarly known as a cost-quality or input-output approach to student performance. He shows that the concept of “equity” is not easily defined; as a result, courts, politicians, school officials, and public debate have, by default, tended to accept “expenditure per pupil” as a measure of equity. Hanushek points out, however, that unless the school system operates efficiently, there is no direct link between expenditures and results.

Hanushek has analyzed the results of 187 prior studies that attempted to relate some objective measure of student output to characteristics of the educational system that were related to costs. He finds that, contrary to conventional wisdom, “the research reveals no strong or systematic relationship between school expenditure and student performance.” His point is not that money doesn't count, but that “unless some way is found to change the districts that would squander additional funds into districts that would use them effectively, added

resources are not likely to lead to any improvement in average performance.”

Hanushek analyzes data from the state of Alabama, which ranks among the lowest in educational expenditures. Alabama has a state-wide Basic Competency Test (BCT), which allows comparisons to be made between school districts on a uniform basis. He estimates the effects of increasing per-pupil expenditures on passing rates for the BCT when controlling for sociodemographic variables (such as family background and rural/urban school district). Bringing the expenditures of all the below-median districts up to the state median level would produce almost no change in the BCT passing rates. Bringing the expenditures of all districts up to the level of the *highest* district in the state (which would bring the state up to about the national mean) would produce at most a 4 percent change in the BCT passing rates.

Hanushek offers a number of additional arguments against simply making policy on the basis of expenditure differences. First, he notes that any effort to lessen variation in expenditure is more likely to increase than to decrease the total level of expenditure. Second, there is no assurance that new funds will go to schools of poor children, since property wealth and concentrations of poverty may exist in the same district. Third, spending differences may not accurately reflect the real resources a district is able to deliver, either because of cost differences for inputs or because of differing needs of student populations. Fourth, districts spend in response to the desires of the population and to population shifts, so their expenditure levels may increase or decrease over time. Fifth, districts perceived to have superior schools will attract home buyers and “bid up” the housing prices in the district relative to otherwise identical housing in another district. Sixth, in many states, issues other than property wealth—such as local preferences, differences in student needs, curricular choices, cost differentials—determine the pattern of expenditure. Seventh, while the tax rate provides an indication of the price that residents pay to raise school funds, differences in tax rates across communities do not necessarily reflect the degree of educational equity.

Hanushek concludes that there is no easy way to improve student performance. He believes that school finance reform that focuses on achieving equal spending per student rather than efficient use of resources will not guarantee improved performance of students. He

admits that no definitive alternative seems sure to bring about improvement in the current school system. However, that is not an argument for maintaining the *status quo*. Rather, Hanushek believes that having some measure of student achievement affect educational expenditures is the route to follow. Hanushek seems optimistic about merit pay for teachers and school choice (such as a voucher system) as potential reforms worth exploring.

Robert Meyer directly addresses the issue of performance measurement in "Can Schools Be Held Accountable for Good Performance? A Critique of Common Educational Performance Indicators." Meyer argues that accurate assessment of student progress is important because administrators and teachers respond to the indications (and resulting incentives) provided by the results of the assessment instrument. He demonstrates that the traditional methods of reporting scores (typically as means or medians of entire school districts) from standardized educational tests may be misleading when used as an assessment of school performance and student achievement. He shows that using such measures as the basis for allocating school system resources (such as merit pay plans for teachers) is not generally desirable, and may even have perverse effects. In particular, these testing methods foster "teaching to the test" at the expense of "real learning."

In agreement with many other critics of large-scale educational testing, Meyer feels that the prevalent pattern of multiple-choice tests that focus on items of factual knowledge rewards rote learning, rather than higher order thinking and the development of problem-solving skills. Since traditional tests are not satisfactory instruments for determining educational achievement, Meyer wants (as do many other critics) performance-based tests that will elicit the student's ability to perform real-world tasks—for example, the road test for a driver's license. As Meyer points out, each student's educational level at a particular moment is the cumulative result of all prior schooling. Therefore, a good measure of educational achievement must not suffer from the three main defects in the traditional assessment measures: nonlocalization, overaggregation across grade levels, and contamination due to mobility. Nonlocalization results from reporting the data from too large an area (such as a school district, or even an entire state). He feels that a good assessment measure must be localized so that it can be related to the school (or even specific classroom) where the learning

occurred—or failed to occur. Overaggregation is the failure to identify what performance gains occurred at a particular grade level. Therefore, performance levels must be measured frequently. Contamination resulting from student mobility occurs when students transfer between school systems. In that case, an assessment score would falsely attribute to one school system the effects (good or bad) of another school system.

Meyer proposes reporting the successes and/or failures of the school system in terms of two related measures of performance: a value added indicator and a gain indicator. The value added indicator would measure *only* the amount a student learned in a particular class in a particular year. This immediately avoids the localization, aggregation, and contamination problems. The gain indicator would report the gain in students' educational level over a period of time (ideally, a school year) *from all sources*. Thus, the sum of all the value added indicators is included, *plus* the contribution of nonschool factors (such as the socioeconomic characteristics of the student's home and community).

Meyer presents the results of several simulations that show the effects of various patterns of student inflow and school effectiveness on student educational gain and achievement. These graphs and tables demonstrate the sensitivity of the apparent measurement of achievement to prior conditions.

In conclusion, Meyer advocates more frequent and better testing, the collection of data on student and family characteristics, and the development of better statistical models.

In "College Choice, Academic Achievement, and Future Earnings," Estelle James and Nabeel Alsalam reinforce the old maxim that the harder one works, the more successful one will be. It should be carefully noted that "success" here is measured only in monetary terms, namely, one's income seven to nine years after graduating from college.

James and Alsalam studied a sample of 1,321 males selected from the National Longitudinal Study of the High School Class of 1972 and the Postsecondary Education Transcript Study data sets. The men were interviewed in 1986, having graduated from college seven to nine years earlier. James and Alsalam used two statistical models to analyze how the characteristics of both the students and the college they attended affected earnings.

In the first model, they compared the 285 colleges that had at least two graduates in the sample. Considering only which college a student graduated from served as a good predictor of that student's future earnings. James and Alsalam then added variables to the first model that represented the initial characteristics of the students (such as high school grades, SAT scores, family income), what they studied, and how well they did in the college. This increased the predictive power of the model, but greatly decreased the importance of the specific college attended.

The second model included the characteristics (such as private/public, large/small, SAT scores of students) of all of the 499 individual colleges represented in the data. This almost eliminated the predictive power of which college the student attended. Including the variables for the characteristics of the students and what they studied and how well they performed academically again increased the predictive power of the model and greatly reduced the importance of the specific college attended.

In sum, James and Alsalam agree with Meyer in that there is both a value added and a gross output (Meyer's "gain") in higher education. What the world perceives—and generally, is willing to pay for—is the "finished product." While there is a component of indirect screening in a particular student's choice of a particular college, what the student does while attending that college—the skills acquired and/or developed—are the best indicator of future financial success.

W. Lee Hansen, in "The Financial Squeeze on Higher Education Institutions and Students: The Balance Between Quality and Access," discusses trends since World War II in the shifting emphasis between the quality of higher education and the ease of access to it. He posits that the goals of the American higher education system are influenced by many forces, both from within and from without.

A noteworthy feature of Hansen's study is that "quality" is measured in terms of financial inputs to education, not academic outputs. Accordingly, "quality" is not the amount students are *learning*, but the amount of *money spent* on faculty salaries and related instructional costs. Correspondingly, "access" here means affordability—whether or not students are able to meet the expenses of attending college.

Hansen calculates monetary proxies that quantify both access and quality. The sum of tuition plus fees, less the total amount of financial

aid, serves as the proxy for access. The sum of all instruction-related costs, adjusted by subtracting the costs of nonteaching programs (such as the extension programs of the land-grant state university systems and sponsored research projects at the major universities) serves as the proxy for quality.

As Hansen observes, scholars from many disciplines have studied patterns of change in society. He particularly refers to previous work on cycles by Arthur Schlesinger. The general tendency seems to be for society to move towards an extreme position, but to then reverse course and head for the opposite extreme, much as the swinging of a pendulum. As evidence for the cycles, Hansen presents tables that show the ebb and flow of the costs of higher education and the sources of student financial support for nine periods from 1947 to 1989.

Analyzing this data, Hansen finds long-run swings between emphasis on access (*e.g.*, the GI bill) and emphasis on quality (*e.g.*, curriculum reforms) in higher education. He finds four of these swings in the last half-century. In the first one, which ran from the late 1930s until the early 1950s, the main thrust was on access to college, with the GI bill providing the ultimate example of this. The next period, from the mid-1950s to the mid-1960s, was an era of enormous expansion in higher education. There was great emphasis on the quality of American higher education, particularly in the areas of engineering and science, brought on by the "space race" with the USSR. In the following period, from the mid-1960s to the beginning of the 1980s, emphasis shifted back to expanding access to higher education, as exemplified by the various federal student aid programs initiated in this period.

The fourth period, which began in the early 1980s and continues to the present, differs from the previous ones. There now appears to be concern about *both* access to and quality of higher education. Hansen concludes that the conflict now is over the "cost-effectiveness" of higher education, and how the costs are to be apportioned among the students and their families, voluntary contributors to higher education, and taxpayers on the local, state, and federal levels.

The most important of the trends that he found is that students and their families are being asked to pay an increasing share of the cost of higher education, rising from 26 percent after World War II to 41 percent by the late 1980s. Hansen offers two plausible explanations for this. One is that the rising demands for other publicly provided ser-

vices (such as welfare and medical care) have increasingly competed with higher education for support from government and private sources. Another possibility is that it increasingly appears to our society that the benefits of higher education (at least in the form of higher incomes) accrue more to the individual than to society; therefore, the greater beneficiary should pay a greater share of the costs.

The last paper in this collection, "The Economics of Education in a World of Change" by Mary Jean Bowman, is quite different from the preceding five. Rather than reporting on research in a particular area of the economics of education, she takes a philosophical approach, exploring more the "what" and "why" of education, in contrast to the "hows" of method and financing of the other papers.

Bowman has a long and distinguished career—indeed, when she began her studies of economics, the subject was much more the qualitative nature of political economy rather than the quantitative path that economics currently follows. Perhaps this makes her more qualified to step back—to see the forest, rather than concentrating on the trees. In trying to predict what kind of education will be most suitable to prepare for a future in which the only certainty seems to be change, it is probably better to take the longest view, rather than try to extrapolate, no matter how carefully, from the trends of the present.

Bowman's central thesis is the importance of exploring change for meaningful analysis in the economics of education. She begins with clarification of the meanings and scope in this context of "education" and "economics of education," going on to specify the kinds of change and the implications of change as they affect and are affected by education in industrialized societies.

Bowman defines education in a broad sense as "learning," which includes but is not limited to "schooling." In considering the very important, and very difficult, problem of how students can best be prepared for the future, she discusses whether general, vocational, or specialized education is most appropriate for coping with the change and uncertain expectations that seem inevitable in the future. She argues for general education, which should provide all students with a solid foundation of both literacy and numeracy. While specific curricula of vocational or specialized education may become obsolete, a sound general education will facilitate lifelong learning—the learning that occurs after graduating from school—which will provide the knowl-

edge needed in one's employment. She says, "a world of change calls for learning both within and outside of schools. It calls for general education in preparation for future learning, for specializations that can cope with change, and for both applied and theoretical learning."

Bowman considers the general learning that should take place in the home and in the preschool years to be of the greatest importance. Not only must children be prepared for formal schooling, but perhaps even more important, children must be socialized in terms of attitudes and behavior in order for them to succeed in the world. She contends that, due to deficiencies in learning in some homes and the perverse trends (such as increased illicit drug abuse) in today's society that work as disincentives to education, schools must, by default, become socializing agencies for many children.

The allocation of resources to educate children for a world that differs in many ways from the one in which we grew up involves many complex issues. While the papers in this volume do not provide answers to all of the problems, it is hoped that the application of an economic perspective will add a useful dimension to the dialogue.

The Economics of Education for At-Risk Students

Henry M. Levin
Stanford University

The nation currently faces an immense crisis in addressing the education of at-risk students—pupils who are unlikely to succeed in existing schools. Such students currently comprise over one-third of all elementary and secondary school students, and their numbers are rising absolutely and proportionately over time. At-risk students are about two years behind grade level in school achievement by sixth grade and perform at about the eighth-grade level if they graduate from high school. Over half do not graduate. Their poor educational performance does not provide them with the skills needed for labor market success and further training, a situation with serious consequences for the economy.

At-risk students are defined as those who are unlikely to succeed in school as these institutions are currently constituted because they do not have the experiences in the home, family, and community on which school success is based. Given the existing curriculum and instructional practices, schools are not neutral arenas in which all types of student backgrounds lead to success. Students who come from middle-class and nonminority backgrounds, with both parents present in their lives, and who speak a standard version of English are much more likely to succeed educationally than those from impoverished, minority, immigrant, nonstandard English-speaking, and single-parent backgrounds. At-risk students are caught in a mismatch between their home situations and what schools require for success. An effective set of policies to improve educational outcomes for at-risk students requires addressing both the in-school and out-of-school experiences of these children.

This article will focus on the contributions that economic analysis can provide in addressing the educational needs of at-risk students. The first part will present information on the demography and educational

status of at-risk students and some economic consequences. The second will offer a summary of what is known about the economic returns to investments in these populations. The final part of the paper will present a new microeconomic approach to the schooling of these youngsters, which has shown promise.

A Crisis of At-Risk Students

The challenge of addressing the needs of at-risk students is important because they are a large and growing portion of student enrollments in the United States, and their poor educational performance has important consequences for the economy and society. It is widely viewed that high school completion represents a minimum qualification for the vast majority of jobs in the U.S. labor force and for eligibility for further training. Students from minority and low-income backgrounds are far more likely to fail to complete high school than other groups, and the proportion of both minorities and children from impoverished circumstances is increasing among the school population.

Among members of the labor force between 25 and 29 years old in 1985, only about 14 percent had failed to complete high school or its equivalent (U.S. Department of Commerce, Bureau of the Census 1987). But the figure among blacks was 19 percent and among Hispanics it was almost 40 percent. Both among minorities and whites, persons from families of low socioeconomic status have considerably higher dropout rates than those from more advantaged backgrounds (Rumberger 1983). Similar patterns exist for academic achievement, in which those from low socioeconomic backgrounds and of minority status show considerably lower test scores than their white and nondisadvantaged counterparts (Smith and O'Day 1991).

The fact that populations of school children who are minorities or from low-income families, especially where the parents have not completed high school, represent a substantial and increasing portion of school enrollments is a particularly ominous situation. From 1970 to 1980, U.S. public school enrollments from the preprimary level to twelfth grade declined from about 46 million to 41 million students.

During the same time period, minority enrollments rose from about 9.5 million to about 11 million, or from about 21 to 27 percent of the total (National Center for Educational Statistics 1984, p. 16). By the year 2020, it is expected that minority children will represent almost half of all children aged 17 and under (Pallas, Natriello, and McDill 1989), a figure that has already been reached in California and Texas. Minority students comprise three-quarters or more of the enrollments of many of the largest cities of the nation, including New York, Chicago, Los Angeles, Philadelphia, Miami (Dade County), and Detroit (McNett 1983). Minority enrollments have been increasing at a more rapid pace than the general population because of considerably higher birth rates and immigration—both legal and undocumented—that have been unprecedented in recent decades. Both factors create rapid growth, particularly among school-age populations. Immigrant and other minority populations tend to be young and of childbearing age, in contrast to an older, nonminority population.

When poverty is used as an indicator for “at-risk” populations, a similar pattern emerges. Between 1969 and 1979 the proportion of children in poverty stayed at about 16 percent; but it rose precipitously to 22 percent by 1983 and is projected to reach 27 percent of the children 17 years and under by 2020 (Koretz and Ventresca 1984; Pallas, Natriello, and McDill 1989). This is a rise from about 15 million to over 20 million children in poverty. Between 1984 and 2020 the number of children who are not living with both parents is expected to rise by 30 percent from 16 million to over 21 million (Pallas, Natriello, and McDill 1989). This is especially alarming, given that the real incomes of single mothers with children fell in absolute terms by 13 percent between 1970 and 1986 (Congressional Budget Office 1988).

Trends for other indicators of children at-risk have been moving in the same direction. For example, Pallas, Natriello, and McDill (1989) project that the number of children raised in families where the mother has not completed high school will rise by 56 percent to over 21 million by 2020. Of particular importance are the low educational attainments of immigrants drawn from rural regions of some of the poorest countries in the world. For example, of the largest single group of immigrants into California—Mexicans—only about 28 percent had more than an eighth-grade education in the early 1980s (Muller 1985, p. 7).

Not only are the numbers of at-risk students growing but there is evidence that their degree of disadvantage is increasing, too. In the fall of 1972 about 46 percent of Hispanic high school graduates participated in postsecondary education immediately following graduation (National Center for Education Statistics 1984, p. 160). By the fall of 1980 that proportion had fallen to 40 percent, despite the widespread loosening of admissions standards during this period. While the participation rate in higher education of Hispanics from middle socioeconomic backgrounds fell by about 10 percent, the rate for Hispanics of lower socioeconomic background fell by 22 percent. This is even more surprising, given that the high school dropout rate for Hispanics rose over the period, meaning that one would normally expect the high school "survivors" to be better qualified. This drastic change in participation over such a short period may have been occasioned by poorer academic preparation and thus lower eligibility for postsecondary education or less adequate financial resources, both factors associated with increasing disadvantage.

In summary, the evidence suggests that the proportion of at-risk students is high and increasing rapidly. Estimates derived from the various demographic analyses suggest that upward of one-third of all students in kindergarten through twelfth grade are educationally disadvantaged or at-risk (Levin 1986). When achievement is used as a criterion, it appears that the number of educationally at-risk students may be as high as 40 percent (Kennedy, Jung, and Orland 1986, pp. 62-63).

General Economic Implications

The rising numbers of at-risk students and their continuing failure to succeed educationally will have important economic ramifications in at least three areas: (1) quality of the entry-level labor force; (2) the cost and quality of higher education; and (3) the cost of public services.

Quality of Entry-Level Labor Force

One consequence of the present educational status of at-risk students will be a serious deterioration in the quality of the labor force. As long as persons from such backgrounds were a small minority of the population, they could be absorbed by low-skill jobs or relegated to the status of unemployment without direct consequences for the economy.

High dropout rates, low test scores, and poor academic performance of a group that will become a larger and larger portion of the school population mean that a larger portion of the future labor force will be undereducated for available jobs. Here we refer not only to managerial, professional, and technical jobs, but even to the lower-level service jobs that are increasingly important in the U.S. economy (Levin and Rumberger 1987). Clerical workers, cashiers, and salesclerks all need basic skills in oral and written communications, the acquisition of which is hardly guaranteed in the schooling of the disadvantaged (National Academy of Sciences 1984). A U.S. government study in 1976 found that while 13 percent of all 17-year-olds were classified as functionally illiterate, the percentages of illiterates among Hispanics and blacks were 56 and 44, respectively (National Assessment of Educational Progress 1976). These and other test score results (Smith and O'Day 1991) suggest that many at-risk students are not acquiring the foundation that will enable them either to work productively in available jobs or benefit from training that would increase productivity and provide job mobility.

As at-risk populations become an increasing and even dominant share of the labor force, their inadequate educational preparation will be visited on the industries and states in which they work, affecting their competitive positions and our national economic status. Employers will suffer in terms of lagging productivity, higher training costs, and competitive disadvantages that will result in lost sales and profits. This problem will be especially severe for states with the largest growth in the disadvantaged population, such as California and Texas, where minorities already represent the majority of all students. It will also be most serious in those industries that depend upon this population for their labor needs. As a result, state and federal governments will suffer a declining tax base and a concomitant loss of tax revenues that could be used to fund improvements in education and other services.

Cost and Quality of Higher Education

The implications for higher education are also severe. Even with high dropout rates, an increasing proportion of high school graduates will come from disadvantaged backgrounds. Without intervention at an early stage in their education, these students will leave high school

with serious learning deficits, which will prevent many of them from benefiting from current levels of instruction in colleges and universities.

High school graduation entitles the at-risk student to pursue postsecondary study in community colleges and many state universities. Even if increasing numbers of disadvantaged students gain college entry, their low achievement means that a high proportion of them will experience academic failure and leave without a degree. Among the group that entered college in 1972, only 13 percent of the Hispanics, 16 percent of the Native Americans, and 24 percent of the blacks completed a bachelor's degree by 1976, compared to 34 percent of the whites (Garibaldi 1986, p. 390). Although ultimate completion rates were higher for all groups, differences remained, and it took longer—on average—for minority students to complete their degrees.

One obvious response to this situation is to provide massive remedial functions to assist educationally disadvantaged students to reach levels where they can benefit from conventional instruction. According to a recent survey by the U.S. Department of Education in the early 1980s, one in every four freshmen was already enrolled in a remedial mathematics course, and one in every six in remedial reading (Abraham 1988). A similar study for fifteen southern states in 1986 found that about 36 percent of the freshmen in public institutions of higher education in those states were taking at least one remedial course in reading, writing, or mathematics (Abraham 1988).

High levels of college failures and dropouts and massive remedial interventions have costly consequences to both students and institutions. Large numbers of failures mean wasted time for students and wasted resources for colleges, not to mention the psychological costs to students of not being able to "make it." Substantial remedial activities require additional faculty, and student programs take longer, with a greater cost in tuition and lost earnings during the extended training period required. Also, as a college or university takes on remedial functions, it is likely to approve some of these courses for degree credit, which results in a watering down of the overall curriculum and standards.

Cost of Public Services

A final consequence of failing to address the challenge of at-risk students will be the rising costs of public services as more and more citizens are forced to rely upon public assistance and undereducated teens, and adults pursue illegal activities to fill idle time and obtain income. Many of the disadvantaged will continue to have difficulty finding regular jobs as adults, so their families will need to depend upon the availability of public assistance to survive. When one applies a teenage unemployment rate of 40 percent or so to a larger and larger group of school dropouts, there are likely to be increasing numbers of undereducated youth taking their activities to the streets rather than to the workplace.

Among a national sample of 19- to 23-year-olds in 1981, 72 percent of the jobless, 79 percent of those on public assistance, and 68 percent of those arrested in the previous year had scored below the average on the AFQT measure of basic skills (Berlin and Sum 1988, p. 29). Among 18- to 23-year-old males in 1981, those with a high school diploma had a 94 percent lower probability of arrest; and among girls aged 18 to 21 the high school graduates had a 54 percent lower probability of having a baby out of wedlock (Berlin and Sum 1988, p. 42).

A study of black women in their mid-thirties in 1982 found that each additional year of schooling was associated with a reduction of about 7 percent in the probability of receiving public assistance (Owens 1990). Moreover, participation in public assistance seems to be becoming even more education-dependent over time; education had twice the impact on the relation in 1982 as it did in 1967 (Owens 1990).

A projection of these outcomes on an expanding at-risk population will not only make the United States a less desirable place to live, but will increase the costs of police services and the criminal justice system. At the same time, the potential decline in economic activity created by an underprepared workforce will erode tax revenues. This situation will place additional pressures on the middle class to pay higher taxes for welfare and the system of criminal justice at the same time that the economy is flagging. As such it will exacerbate the political conflict between haves and have-nots, as taxpayers resist raising

taxes in the light of a faltering economy and mounting pressures for higher expenditures.

Summary of General Economic Implications

To fail to address the present and future educational needs of at-risk students will incur high social costs in terms of reduced productivity in the labor force and higher education as well as rising costs of public services. Education is not only linked to public assistance and criminal justice, it is also linked to health, status, and a variety of other important social outcomes (Haveman and Wolfe 1984). In fact, when all the identifiable outcomes associated with education are taken into account, it has been estimated that the overall return on education is twice as high as when only its effect on income is considered.

Benefit-cost Studies of Educational Investment

The knowledge that economic and social benefits can be achieved by investing in at-risk student populations is not an adequate criterion for investment. Although such investments are likely to result in considerable benefits, there are also likely to be considerable costs. From an economic perspective, it is necessary to know whether benefits exceed costs and whether they exceed them by magnitudes equal to or greater than alternative social investments. In this section, we will review the results of benefit-cost studies of educational investments among those populations.

Programs for Reducing High School Dropout Rates

A number of economic studies have addressed the costs and benefits of programs for reducing the rate of high school dropouts. In a classic study on the subject, Weisbrod compared the impact of a St. Louis program designed to reduce the rate of dropouts among “dropout-prone” high school students with the rate of dropouts in a control group of similar students who did not have such a program (Weisbrod 1965). The dropout prevention program was associated with a high school completion rate that was about 7 percent higher than that of the control

group. Weisbrod estimated the cost for each of the additional graduates and contrasted it with the estimated income benefits of high school graduation for these students. He found that the costs of the program exceeded its benefits.

There are at least two reasons for believing that analyses of more recent programs would show stronger benefits. Weisbrod used 1959 census data to estimate the additional incomes of the graduates. Because of discrimination and other factors, the earnings of women and minorities were a much smaller portion of white male earnings some 30 years ago than they are today. Since the dropout-prone group included considerable numbers of females and minorities, the benefits were probably considerably understated relative to what would be obtained with more recent data. Further, the earnings advantages of high school graduates relative to dropouts have increased. Finally, the program that Weisbrod evaluated was initiated over thirty years ago when dropout prevention was in its infancy.

In contrast, a more recent study of dropout prevention found large net benefits (Stern, Dayton, Paik, and Weisberg 1989). This evaluation was based upon the success in reducing the number of dropouts at eleven academies created in public high schools in California. These academies comprised special programs or schools within the larger high school setting and provided vocational training for careers in which students stood a good chance of placement, as well as academic training. Students were given special attention from their teachers and the representatives of local employers. When students were matched with a similar group of students in regular school programs, it was estimated that the academies had saved 29 persons who would have been expected to drop out.

The marginal costs of the academy program, beyond those of the regular school program for all 327 students, were compared to benefits in terms of the additional earnings of the twenty-nine persons "saved" from dropping out. The overall benefits of the program were found to exceed overall costs by considerable amounts, the specifics depending upon which assumptions were used regarding benefits. However, the results also show that for some of the academies net benefits were positive and for others negative—that is, costs exceeded benefits. This suggests that a more refined evaluation of individual programs would

be useful in arriving at an understanding of which programs were the most promising on the basis of a benefit-cost analysis.

In contrast to studies of a single dropout program, Levin undertook a national study on the economic consequences of high school dropouts (Levin 1972). Here he calculated the additional lifetime earnings and tax revenues that would have been generated if the entire cohort of 25- to 34-year-old males in 1970 had graduated from high school. It was assumed that even if existing dropouts had graduated they would *not* have done as well as those who had actually graduated from high school. Thus, additional earnings of dropouts who would be induced to graduate were assumed to be only 75 percent of those of conventional high school graduates. But it was also assumed that a portion of the induced graduates would continue into higher education, with resulting additional earnings from that source as well.

The total loss of lifetime earnings for this group as a result of failure to complete at least high school was estimated at about \$237 billion. The additional cost for achieving this result was comprised of two parts: first, the cost of the additional years of schooling undertaken by members of the group; second, the cost of additional expenditures to prevent dropping out. It was assumed that it would have been necessary to increase annual schooling expenditures on those at-risk of dropping out by 50 percent a year for all of their elementary and secondary schooling to keep them in school until completion of high school. On this basis, it was estimated that the total costs of achieving at least high school graduation for all members of the cohort was about \$40 billion, producing a benefit of \$6.00 for each dollar of cost. The additional lifetime earnings would have generated about \$71 billion in government revenue or about \$1.75 in tax revenues for each dollar in cost. The study also estimated that inadequate education was contributing about \$6 billion a year to the costs of welfare and crime in 1970.

Robledo (1986) replicated this analysis more recently for that cohort of Texan ninth graders in 1982–83 who were projected to drop out before their anticipated graduation in 1986. They estimated the benefits of a dropout prevention program as those attributable to savings in public assistance, training and adult education, crime and incarceration, unemployment insurance and job placement, and as higher earnings associated with the additional number of high school graduates. Such benefits were calculated at \$17.5 billion, and the costs to elimi-

nate dropouts for this cohort were estimated at slightly less than \$2 billion or a ratio of \$9 in benefits for each dollar of costs. Estimates of additional tax revenues were 2.5 times greater than costs to the taxpayer.

Catterall (1987) did a similar type of analysis for persons who dropped out of the Los Angeles high school class of 1985. He found that because of high school dropouts, the Los Angeles class of 1985 was projected to generate over \$3 billion *less* in lifetime economic activity than if all of its members had graduated. In contrast, Catterall suggested that the cost of investing successfully in dropout reduction would be a mere fraction of this amount. Further, he found that Los Angeles was addressing the dropout problem with specific programs that were spending the equivalent of only about \$50 per dropout, or less than one-half of 1 percent of school spending, even though 40 percent of its students were not graduating.

Preschool and Higher Education

There is evidence that even preschool investments in at-risk populations can reduce dropping out as well as provide other types of benefits. Barnett undertook a benefit-cost analysis of the Perry Preschool Project in Ypsilanti, Michigan (Barnett 1985). The Perry Preschool approach has been studied for two decades and has been used as a model for hundreds of preschools for disadvantaged students across the country, including the national Head Start program. Students who had been enrolled in the preschool project were followed until age 19. It was found that relative to a matched control group, enrollees in the project experienced better school achievement, educational placement, educational attainment, and employment. Monetary values for the benefits were calculated on the basis of the apparent effect of these advantages on the value of childcare during the programs; reduced school expenditures for remediation, special services, and grade repetition; reduced costs of crime, delinquency, and welfare; and higher earnings and employment.

It was found that the benefits exceeded the costs by a large margin under a wide range of assumptions. The one-year program showed benefits of \$7.00 for every dollar of costs, a benefit-cost ratio of about 7:1, and the two-year program showed a benefit-cost ratio of about

3.6:1 (Berrueta-Clement et al. 1984, p. 60). About 80 percent of the net benefits were received by taxpayers in the form of higher tax contributions and lower expenditures on education, crime, and welfare and by potential crime victims in the form of lower costs for property losses and injuries.

A study of benefits and costs for financial aid to stimulate participation in higher education for low-income students has also indicated high benefits relative to costs for government investment (St. John and Masten 1990). Here researchers compared tax revenues generated by the additional income produced by the higher levels of college participation among low-income students with the costs of financial aid that induced these higher enrollments. The net present value of additional tax revenues was four times as great as the cost of the aid program for students in the high school class of 1980. That is, from the perspective of the federal treasury, such programs had a benefit-cost ratio of 4:1.

These particular studies suggest that investments in at-risk students yield high returns to society. Such social investments are highly worthwhile in that their benefits exceed costs and that the margin by which they exceed costs is competitive with or superior to that of other highly productive investments. Of greatest importance is that higher tax revenues and reductions in the costs of social services more than compensate for the investments. In fact, in the case of the early childhood intervention program established by the Perry Preschool, most of the net benefits accrued to taxpayers (Barnett 1985).

Summary of Benefit-Cost Results

These benefit-cost results suggest that investments in the education of students at risk of undereducation are likely to have high payoffs to society. While each study can be questioned because of imperfect information and the need to make assumptions on both the cost and benefit sides of the equation, their overall pattern is remarkably consistent. This interpretation is buttressed by a recent study that found that increased investment in schooling quality among states was consistently associated with higher earnings of the adults who were schooled in those states, holding constant other influences (Card and Krueger 1992).

Estimated benefits for educational interventions tend to be about three to six times as high as estimated costs for at-risk students. According to Haveman and Wolfe (1984), the consideration of returns to human capital investments in the form of increases in earnings will capture only about half of the total returns. Thus most of these estimates are subject to understatement because they tend to be limited to the effects of educational investments on productivity and earnings and do not capture the value of reductions in the costs of health, public assistance, criminal justice, and a variety of other benefits. However, recent work suggests that cross-sectional estimates tend to overstate the benefits to human capital investments on behalf of the poor (Levin and Kelley 1991). All of the estimates are based upon cross-sectional evidence, with the exception of those based upon the preschool intervention. Since there is no direct evidence on the potential degree of overstatement or understatement of these results, a reasonable assumption is that they are offsetting and that the estimates are a reasonable first approximation of returns to investments on behalf of at-risk populations.

The Microeconomics of Educational Reform

In the early 1980s, a rash of reports by national commissions and other groups were published recommending national educational reforms to improve economic competitiveness. The most important of these was *Nation at Risk*, produced by the National Commission on Excellence in Education (1983). Most of the recommendations of these reports addressed changes in secondary school programs for college-bound students by calling for more academic courses with more rigorous standards at that level. But at-risk students were not even meeting the "lower" standards that existed at that time and were dropping out in response to academic demands. The reports said almost nothing about improving school effectiveness prior to high school to make it possible for at-risk students to meet both existing and higher standards.

Why were the reports of these commissions silent about at-risk students? In response to this question, I undertook a study on the demography, educational outcomes, and social consequences of this group of

students (Levin 1986), the results of which are summarized in the first section of this paper. As an extension of that study I began to explore the production of schooling for this group of children (Levin 1988). Surprisingly, I found that the educational process in schools attended by these children was the *cause* of much of the problem rather than the solution.

That research found that at-risk students started behind other students and lagged farther behind the educational mainstream the longer that they were in school. And this problem did not appear to stem from a lack of teacher dedication, a charge that has often been made. Paradoxically, it occurred because compensatory programs for the disadvantaged are designed to slow down the instruction of such students, on the that assumption that at-risk students are less capable than others. Such students are placed into less demanding instructional settings—either by pulling them out of their regular classrooms or by adapting the regular classroom to their “needs”—and offering remedial or compensatory educational services. While this approach appears to be both rational and compassionate, it has exactly the opposite consequences.

First, it reduces learning expectations on the part of both the children and the educators assigned to teach them, and it stigmatizes both groups with a label of inferiority. Second, it slows down the learning process so that at-risk students fall farther and farther behind the mainstream, the longer that they are in school. Third, the approach to remediation is to provide repetitive practice of low-level basic exercises through endless drill and practice. This educational experience is empty and joyless because it fails to incorporate a rich curriculum, student involvement and discourse, interesting applications of concepts, active problem solving, and learning activities that build on the strengths of the students and their backgrounds. Finally, this remedial approach does not draw sufficiently upon parental and community resources, nor does it provide for the participation of school-based educators to influence the programs that they must implement.

The study concluded that an effective approach to educating the disadvantaged must be characterized by high expectations, deadlines by which such children will be performing in the educational mainstream, stimulating instructional programs, planning by the educational staff who will offer the program, and the use of all available resources, including the parents of the students. This approach should incorporate

a comprehensive set of strategies that mutually reinforce each other in creating an organizational push toward raising the achievement of students to the level that we expect in the mainstream.

A key element in this strategy is accelerated schools, which were designed by our Stanford Accelerated Schools Project to have exactly the opposite consequences by bringing at-risk students into the educational mainstream by the end of elementary school (Levin 1988). Our premise was very basic: at-risk students must learn at a *faster* rate than more privileged students—not at a slower rate that drags them farther and farther behind. What is required is an enrichment strategy rather than a remedial one.

I hypothesize that acceleration works as well for at-risk students as it has for their better prepared counterparts. One recent study assigned at-risk students at random to remedial, average, and honors classes in seventh-grade mathematics. At the end of the year, the at-risk students in the honors class—which provided pre-algebra instruction—outshone at-risk students in the other two groups (Peterson 1989). Similar results were found when at-risk students were provided with high-content instruction that emphasized thinking ability and decision making rather than basic skills (Knapp, Shields, and Turnbull 1992).

Institutionalizing Change

Moving from an idea to institutional change is never an easy process. In order to develop a strategy for creating accelerated institutions, we found that we would have to make three major changes in U.S. schools, changes that were in deep conflict with current practices (Levin 1988). These changes have deep economic roots in that they require that: a clear objective function for the school (unity of purpose) be established; those with *de facto* property rights exercise those rights on behalf of children within a framework of incentives and accountability (school-site empowerment with responsibility); and an appropriate technology of schooling that will deliver results (building on strengths) be employed.

Unity of Purpose

Most schools that educate at-risk students seem to lack any central purpose. In economic terms they are firms without an objective func-

tion. In this framework, traditional schools are better understood as a composite of individuals and programs that seem largely disparate and piecemeal with no central vision. Planning, implementation, and evaluation are typically done independently and by different groups. Teachers tend to see their responsibilities extending no farther than maintaining good practices in self-contained classrooms, while remedial specialists work in isolation from each other and the regular school program.

Acceleration requires the establishment and pursuit of a common vision that serves as a focal point for the efforts of parents, teachers, staff, and students. The vision of an accelerated school must focus on bringing children into the mainstream, where they can more fully benefit from school experiences and opportunities. The development of this vision requires the combined efforts and commitment of all parties involved. Unity of purpose refers to both a vision or dream of what the school can be and an action plan that will get the school there.

School-site Empowerment

Existing schools for at-risk students are largely dominated by decisions made by entities far removed from the school site and classroom. Federal and state governments and central offices of school districts have established a compendium of rules, regulations, directives, policies, laws, guidelines, reporting requirements, and "approved" instructional materials that serve to stifle educational decisions and initiative at local school sites. It is little wonder that administrators, teachers, parents, and students tend to blame factors "beyond their control" for the poor educational outcomes of at-risk students. And, as the historical record has shown, compliance with these policies ensures failure, not success.

Accelerated schools are based on the concept of internal responsibility, in which major decisions that will determine educational outcomes are made by establishing a collective sense of efficacy and applying the skills and organization to undertake the changes that are necessary. If the school is to achieve its vision of educational success, administrators, teachers, other staff, parents, and students must participate in making informed decisions regarding school activities. Important areas of school-site decisions include some or all of the following: curriculum, instructional strategies, instructional materials, personnel, and

allocation of resources inside of the school. Such decision making requires active support from the district's central office in the form of information, technical assistance, staff development, and evaluation, as well as an overall system of accountability in which the school is rewarded according to its performance.

Building on Strengths

Schools with large numbers of at-risk students tend to highlight the weaknesses of their students, staff, funding, administrative support, and so on, as an explanation for poor performance. A particularly heavy emphasis is placed on the litany of what is wrong with at-risk students and their parents. But good pedagogy begins with the strengths and experiences of participants and builds on those strengths rather than dwelling on the weaknesses. This means that schools must shift from a technology of production that has shown consistent failure to one that has shown superior results.

Accelerated schools seek out the strengths of their students and other participants and use those strengths as foundations on which to build their programs. In this respect, students are treated as gifted and talented students, where strengths are identified which are then used as a basis for providing enrichment and acceleration. The strengths of at-risk students are often overlooked because they are not as obvious as those of middle-class students. But our research has shown that at-risk children bring assets that can be used to accelerate the learning process. These include interest and curiosity in oral and artistic expression, ability to learn through manipulation of appropriate learning materials and interesting applications, the capability to delve eagerly into intrinsically interesting tasks, and a capacity for learning to write prior to mastering reading skills.

The process of building on strengths is not limited to students. Accelerated schools also build on the strengths of parents, teachers, and other school staff. Parents and teachers are largely underutilized resources in most schools. Because they want their children to succeed, parents can be powerful allies if they are placed in productive roles and provided with the skills to work with their children. Teachers bring gifts of insight, intuition, and organizational acumen to the instructional process, gifts often untapped by the mechanical curricula so typical of remedial programs. Accelerated schools acknowledge the gifts

of teachers and parents and build on those strengths in fulfilling their accelerated visions.

Combining the Principles

An accelerated school is not just a conventional school with new principles or special programs grafted onto it. It is a dynamic environment in which the entire school and its operations are transformed. The emphasis is on the school as a whole, rather than on a particular grade, curriculum, staff development approach, or other limited strategies. The goal is high academic achievement for *all* students.

The three principles of unity of purpose, site-based empowerment, and building on strengths are woven together in virtually all the activities of the accelerated school. The school is governed by its staff, students, and parents, and priorities are pursued by task groups that follow a systematic inquiry process for problem solving, implementation, and evaluation.

Accelerated schools use a heavily language-based approach across all subjects, even mathematics, with an early introduction to writing and reading for meaning. Curricula reflect a sense of high expectations and a tie to the students' cultures. Active learning experiences are provided through independent projects, problem solving, and utilizing new knowledge and skills in concrete situations. By applying academic concepts and skills to real-life problems and events, students see the usefulness of what they are learning.

The organization of accelerated schools allows for a broad range of participants and a collaborative approach in which students' families play a central role. Indeed, success depends on parents working with staff and students, helping to make school decisions by participating in the decision bodies of the school.

Some Results of Accelerated Schools

The first two accelerated pilot schools were established in 1987 and have been operating for five years. The total transition from a traditional to an accelerated school takes about six years. Since that time approximately three hundred additional schools, most of which are elementary schools, with a recent extension to middle schools, have initi-

ated the transition process. We have found that the transformation to an accelerated school can be done primarily by reallocating existing resources to free up staff time and make other provisions for staff development and accelerated school activities. To my knowledge, none of these schools has obtained additional funding beyond even 1 percent of their budgets to pursue accelerated school activities. We believe that the basic transformation to and operation of an accelerated school can be done largely within existing resources. It should be noted that most of the other national educational reforms that have shown success require an additional cost of about \$1,000 per student, in comparison with about \$20 to \$30 per student for accelerated schools.

Early results have been extremely promising. The Daniel Webster School in San Francisco enrolls a student body that is over 90 percent minority and over 80 percent on public assistance. It was one of the bottom elementary schools in San Francisco in 1987, ranking sixty-fifth out of sixty-nine schools with test scores in mathematics. By 1991 the mathematics scores had risen to twenty-third in San Francisco, among the top third of all schools. Students were performing above grade level in mathematics at every grade. Test score gains in all three areas tested—reading, language, and mathematics—were the highest of all the schools in San Francisco. The Daniel Webster School was the only school in San Francisco in which both black and Spanish-surname students made more than a year of academic progress in one academic year.

The Hollibrook Elementary School in Houston enrolls over one thousand students, many of them recently arrived immigrants from Central and South America. About 90 percent of the students are from families below the poverty line. In 1988 the school's fifth graders were about two years behind grade level in reading and language arts and almost half a year behind grade level in mathematics. By the spring of 1991 Hollibrook fifth graders were performing at grade level in all subjects and one year above grade level in mathematics (McCarthy and Still 1993).

Most of the accelerated schools have been established in the last two years, so it is too early for them to have completed their transformation. Nevertheless, the early results for these schools are also impressive, with improved attendance, parent participation, test scores, student projects, and reduced behavior problems and vandalism.

Investment in the education of at-risk students has a large payoff, and we have the wherewithal to use that investment wisely in accelerated schools. Indeed, those characteristics that make for an efficient firm can be applied to schools to improve their efficiency substantially. Given this evidence, it is surprising that many economists immediately resort to a market approach in looking for economic strategies to improve the education of at-risk students (Friedman 1962; Levin 1991). Typically, they cite the work of Chubb and Moe (1990) or Hoffer, Greeley, and Coleman (1987), which was reanalyzed by Willms (1987), who found that students in Catholic schools were able to achieve as much as a one-tenth of a standard deviation advantage over similar students in public schools. But accelerated schools have shown achievement gains of 1.5 standard deviations, or fifteen times that large, without resorting to a change in educational finance to vouchers or other systems that would require public funds for private schools. No comparison between private and public schools has come close to finding this effect.

References

- Abraham Jr., Ansley A. 1988. "Remedial Education in College: How Widespread Is It?" *Issues in Higher Education*, No. 24. Atlanta: Southern Regional Education Board.
- Barnett, W. Steven. 1985. "Benefit-Cost Analysis of the Perry Preschool Program and its Long-Term Effects," *Educational Evaluation and Policy Analysis* 3: 333-42.
- Berlin, Gordon, and Andrew Sum. 1988. "Toward a More Perfect Union: Basic Skills, Poor Families." Occasional Paper 3. Ford Foundation Project on Social Welfare and the American Future, New York.
- Berrueta-Clement, John R., Laurence J. Schweinhart, W. Steven Barnett, Ann S. Epstein, and David P. Weikart. 1984. "Changed Lives: The Effects of the Perry Preschool Program on Youths Through Age 19." Monograph No. 8, High/Scope, Ypsilanti, MI.
- Card, David and Alan Krueger. 1992. "Does School Quality Matter? Returns to Education and the Characteristics of Public Schools in the United States," *Journal of Political Economy* 100, 1 (February): 1-40.

- Catterall, James S. 1987. "On the Social Costs of Dropping Out of School" *High School Journal* (October-November): 19-30.
- Chubb, John E. and Terry M. Moe. 1990. *Politics, Markets and America's Schools*. Washington, DC: Brookings Institution.
- Congressional Budget Office, United States Congress. 1988. *Trends in Family Income: 1970-1986*. Washington, DC: Government Printing Office.
- Friedman, Milton. 1962. "The Role of the Government in Education." In *Capitalism and Freedom*. Chicago: University of Chicago Press, Chap. 6.
- Garibaldi, Antoine. 1986. "Sustaining Black Educational Progress: Challenges for the 1990s," *Journal of Negro Education* 55, 3.
- Haveman, Robert and Roberta Wolfe. 1984. "Schooling and Economic Well-Being: The Role of Nonmarket Effects," *Journal of Human Resources* 19, 3 (Summer): 377-407.
- Hoffer, Thomas, Andrew Greeley, and James S. Coleman. 1987. "Catholic High School Effects on Achievement Growth." In *Comparing Public and Private Schools*, edited by Edward H. Haertel, Thomas James, and Henry M. Levin. Vol. 2. *School Achievement*. New York: Falmer Press, pp. 67-88.
- Kennedy, Mary M., Richard K. Jung, and Martin E. Orland. 1986. *Poverty, Achievement, and the Distribution of Compensatory Education Services*. Interim Report from the National Assessment of Chapter 1. U.S. Department of Education, Office of Educational Research and Improvement, Washington, D.C.
- Knapp, Michael S., Patrick M. Shields, and Brenda J. Turnbull. 1992. *Academic Challenge for the Children of Poverty*. Vols. 1&2. Washington, DC: U.S. Department of Education. Office of Policy and Planning.
- Koretz, Daniel and Marc Ventresca. 1984. *Poverty Among Children*. Washington, DC: Congressional Budget Office.
- Levin, Henry M. 1972. "The Costs to the Nation of Inadequate Education." Report prepared for the Select Senate Committee on Equal Educational Opportunity, 92d Congress.
- . 1986. *Educational Reform for Disadvantaged Students: An Emerging Crisis*. West Haven, CN: NEA Professional Library.
- . 1988. "Accelerating Elementary Education for Disadvantaged Students." In *School Success for Students at Risk*. Council of Chief State School Officers. Orlando, FL: Harcourt Brace Jovanovich, pp. 209- 226.
- . "The Economics of Educational Reform." *Economics of Education Review* 10, 2: 137-158.
- Levin, Henry M., and Carolyn Kelley. 1991. "Can Education Do It Alone?" Paper presented at Sixty-Fifth Annual Conference of the Southern Economic Association, Nashville (November 25).

- Levin, Henry M., and Russell W. Rumberger. 1987. "Educational Requirements for New Technologies: Visions, Possibilities and Current Realities," *Educational Policy* 3: 333-354.
- McCarthy, Jane, and Suzanne Still. 1993. "Hollibrook Accelerated Elementary School." In *Restructuring School Learning from Ongoing Efforts*, edited by Joseph Murphy and Phillip Hallinger. Newbury Park, CA: Corwin Press.
- McNett, Ian. 1983. *Demographic Imperatives for Education Policy*. Washington, DC: American Council on Education.
- Muller, Thomas. 1985. *The Fourth Wave: California's Newest Immigrants*. Washington, DC: Urban Institute Press.
- National Academy of Sciences. 1984. "High Schools and the Changing Workplace." Report of the Panel on Secondary School Education for the Changing Workplace. Washington, DC: National Academy Press.
- National Assessment of Educational Progress. 1976. *Functional Literacy and Basic Reading Performance*. Washington, DC: U.S. Office of Education. Department of Health, Education, and Welfare.
- National Center for Educational Statistics. U.S. Department of Education. 1984. *The Condition of Education*, 1984 Edition. Washington, DC: Government Printing Office.
- National Commission on Excellence in Education. 1983. *Nation At Risk*. Washington, DC: Government Printing Office.
- Owens, John D. 1990. "The Social Benefits of Education: An Intertemporal Analysis." Economics Department, Wayne State University.
- Pallas, Aaron M., Gary Natriello, and Edward L. McDill. 1989. "The Changing Nature of the Disadvantaged Population: Current Dimensions and Future Trends," *Educational Researcher* 5 (June-July): 16-22.
- Peterson, John M. 1989. "Remediation is No Remedy," *Education Leadership* 46, 6: 24-25.
- Robledo, Maria del Refugio, ed. 1986. *Texas School Dropout Survey Project: A Summary of Findings*. San Antonio: Intercultural Development Research Association.
- Rumberger, Russell. 1983. "Dropping Out of School: The Influences of Race, Sex, and Family Background," *American Educational Research Journal* 20: 199-220.
- St. John, Edward P., and Charles L. Masten. 1990. "Return on the Federal Investment in Student Financial Aid: An Assessment for the High School Class of 1972." Department of Educational Leadership, University of New Orleans. Typescript.

- Smith, Marshall, and Jennifer O'Day. 1991. "Educational Equality: 1966 and Now." In *Spheres of Justice in Education*, edited by Deborah A. Verstegen and James G. Ward. New York: Harper Business, pp. 53–100.
- Stern, David, Charles Dayton, Il-Woo Paik, and Alan Weisberg. 1989. "Benefits and Costs of Dropout Prevention in a High School Program Combining Academic and Vocational Education: Third Year Results from Replications of the California Peninsula Academies," *Educational Evaluation and Policy Analysis* 11, 4 (Winter): 405–416.
- U.S. Bureau of the Census. 1987. "Educational Attainment in the United States: March 1982 to 1985." *Current Population Reports*, ser. P-20, no. 415, Washington, DC: Government Printing Office.
- Weisbrod, Burton A. 1965. "Preventing High School Dropouts." In *Measuring Benefits of Government Investments*, edited by Robert Dorfman. Washington, DC: Brookings Institution, pp. 117–148.
- Willms, J. Douglas. 1987. "Patterns of Academic Achievement in Public and Private Schools: Implications for Public Policy and Future Research." In *Comparing Public and Private Schools*, edited by Edward H. Haertel, Thomas James, and Henry M. Levin. Vol. 2. *School Achievement* New York: Falmer Press, pp. 113–134.

Can Equity Be Separated from Efficiency in School Finance Debates?

Eric A. Hanushek^{*}
University of Rochester

School finance discussions have concentrated on equity and rest on a few elementary premises. Poor children, often residing in decaying cities, do worse in terms of achievement, jobs, and overall success than children from better environments. Schools are society's designated institution to remedy this situation—but schools serving the disadvantaged are hampered in this task by a lack of sufficient resources. With more funding, these schools could put in place the successful programs that are available, and the cycle of poverty could be broken. It is only equitable then to support poor schools at the level at which schools for otherwise more advantaged students are financed. The more recent variant of the discussion, focusing on an adequacy version of equity, begins by noting the need for high quality education in order for an individual to compete successfully in the labor market, and then turns to a statement of how overall funding for schools must be increased to provide everybody with acceptable opportunities.

These common arguments are simple, straightforward, and compelling. Unfortunately, they are also seriously flawed. The quest for equity has generally pointed to policies that neither promote greater equity nor help deal with the serious schooling problems facing the United States.

For over two decades, courts and legislatures have been embroiled in debate and controversy over the way in which local public schools are financed. Interestingly, this has been an area where the states have completely dominated policy deliberations, and the federal government has never played an important role. Indeed, as a direct result of the U.S. Supreme Court ruling in *San Antonio Independent School Dis-*

---

^{*}This is a revised and expanded version of Hanushek 1991.

trict v. Rodriguez, the court discussion has been conducted exclusively at the state level. Each state has followed a different course based on the requirements of its state constitution, the preferences of its citizens and legislators, and the wisdom of its courts. Nevertheless, while sometimes obscured by the details of specific state actions, there are common elements to the school finance policy developments in the states.

One important lesson learned over time is that school finance court cases, legislative decisions, and school policies in general are more complicated than was previously thought. The framework for deliberations on school finance reform was developed in the 1960s and given national attention through the landmark case in California, *Serrano v. Priest*. This case, which has been transported elsewhere, set out what is now the standard argument:¹ (1) Traditional funding of schools, which relies heavily on local funds raised substantially by property taxes, leads to large disparities in the education available to rich (suburban) students and to poor (urban and rural) students; and (2) The inequities in the quality of schooling resulting from the fiscal system must be corrected, and the courts are an obvious route to forcing the legislature to provide the economically and educationally disadvantaged with better schools.

An updated version of these arguments is found in Kozol (1991), where the disparities in schools between some of the nation's best and worst schools are described in vivid detail. Armed with this descriptive information, Kozol proceeds directly to the policy conclusion that all schools should be moved to duplicate the very best, a conclusion that merges both the equity and the adequacy arguments.

We have now discovered that many of these simplistic views are misleading, if not just plain wrong. The required actions involve more fundamental adjustments than merely redirecting funds, and these fundamental changes are difficult to implement directly from the court or from the capitol. These complications are addressed in the subsequent discussion.

This paper considers the overarching public policy issues involved in searching for improved equity through altering school financing arrangements, concentrating on the central policy issues that transcend state boundaries. Moreover, it avoids all consideration of legal theories

and interpretations that have surrounded the court cases except as they intersect with larger matters of broader educational policy.

Most school finance discussion, as opposed to school policy discussion, has focused almost exclusively on variations in expenditure per student. A variety of reasons can be cited to explain this emphasis. First, expenditure levels are easily measured and easily modified by the court or legislature. Second, it seems reasonable to presume that what a school can offer in terms of quality or breadth is directly correlated with the resources devoted to the task. Finally, even if there is some doubt about how well money is currently being spent, money well spent would surely make a difference. Each of these premises is reasonable *if schools are operating efficiently*.² Given efficient school operation, expenditure is a good index of performance. On the other hand, if schools are not operating efficiently, the interpretation of expenditure differences becomes totally ambiguous, because expenditure variations need no longer be directly correlated with variations in school quality. Moreover, added funding of schools may lead to no gains in student performance.

One fundamental observation underlies the discussion in this paper: *There is no systematic relationship between school expenditure and student performance*. This observation implies a significant level of inefficiency in schools. Given that, legal arguments and policy decisions based on expenditure variations are simply suspect, at least from an educational perspective. Indeed, many popular changes, both proposed and adopted, no longer look like "reform" but instead tend to move us away from good policy.

School finance discussions have not totally ignored the potential pitfalls of concentrating on expenditure alone. After passing references to issues of efficiency along with assertions that the research is ambiguous,³ pragmatism is frequently claimed as the underlying justification: expenditure differences appear to be such a reasonable measure of differences in schools, and they are measurable.⁴ I argue later that this logic is likely to cause serious distortion in policies.

The plan of this paper is straightforward. It begins with a discussion of the evidence about expenditure and school performance. It then considers how this evidence relates to court cases and overall judgments about a state's schools. It concludes with an examination of how court

cases, and the related legislative actions, relate to effective policies toward schools.

What We Know About School Expenditure

Because the interpretation of expenditure differences is so central to all discussions of school finance, this section provides evidence concerning the relationship between expenditure and student performance.⁵ It is simply not possible to ignore these data in setting school policy when the objective is either to improve overall student performance or advance the cause of true educational equity.

Aggregate Data

Much of the current concern about the performance of our schools is motivated by the fact that student performance has remained constant or actually fallen during a period in which school spending has continually increased. Figure 1 illustrates this by superimposing the trend in student performance on the trend in educational expenditure. Real expenditure per pupil has risen steadily and dramatically over the past two decades. Specifically, after allowing for inflation, expenditure per pupil more than doubled between 1967 and 1991; this corresponds to about a 3.5 percent compound annual growth rate.⁶ At the same time, performance as measured by Scholastic Aptitude Test (SAT) scores fell to a level significantly below those attained during the mid-1960s. Moreover, while there was some recovery from the 1979-80 trough, the improvements of the early 1980s have now ceased.

There are reasons for quibbling about these specific statistics for both achievement and spending. The measurement of performance by SAT scores has been questioned because the test does not rely on a representative sample, because the test-taking population has changed over time, and because the content of the test itself may have changed. Analysis of these objections, however, indicates clearly that the observed achievement decline is not simply an artifact of that specific test. Further, declines have been registered on a variety of other tests given over the same time (see Congressional Budget Office 1986,

1987). Continued international evidence also places U.S. students behind a surprisingly wide range of foreign students on math and science performance. For example, in tests of advanced algebra for twelfth graders in 1982, U.S. students trail students from Hong Kong to Hungary, bettering only the students from Thailand in fifteen countries sampled (McKnight et al. 1987).⁷ Thus, there is no doubt that students are performing worse now than they did in the past, when spending on schools was noticeably less.

Figure 1
Real School Expenditure and Achievement
1967–1991

*Current expenditure in 1992 dollars per student in average daily attendance (ADA).

Similarly, some have argued that the tasks facing schools have changed over time so that the comparisons of expenditures are not strictly appropriate. For example, increased expenditure may partly

reflect attempts to educate more expensive students—handicapped students, immigrants, and other educationally disadvantaged. Again, however, while these changes in student populations undoubtedly have some influence on costs, they are insufficient to explain the substantial aggregate increases that have transpired.

Moreover, it is important to note that the expenditure patterns reflect a number of underlying adjustments, which mirror common policy recommendations. Pupil-teacher ratios have fallen steadily for the past three decades. While there were twenty-five students per teacher in public elementary and secondary schools in 1965, there were fewer than eighteen in 1985.⁸ Over the same period, the proportion of teachers holding a master's degree or above went from under a quarter to over half. Median teacher experience also almost doubled, going from eight years in 1966 to fifteen in 1986.⁹ The only aggregate input not to follow this steady pattern is teacher salaries. Real teacher salaries, as best we can tell, have cycled: average salaries rose through the 1960s, fell back in the mid to late 1970s, and rose again during the 1980s.¹⁰

The aggregate picture is clear. School spending has increased dramatically since the mid-1960s, largely through the instituting of policies that educational decision makers have proposed as a way of improving student performance—reducing class sizes and upgrading the education and experience of the teaching force. Yet student performance has actually fallen over the same period.

Individual- and School-Level Analyses

Although research into the determinants of students' achievement takes various approaches, one of the most appealing and useful is what economists call the *production function* approach, or in other disciplines the *input-output* or *cost-quality* approach. In this approach, attention is focused primarily on the relationship between school outcomes and measurable inputs into the educational process.

The origin of estimating input-output relations in schools is usually traced to the monumental U.S. study, *Equality of Educational Opportunity*, or what is more commonly known as the Coleman Report. Explicitly designed to study equity, this report was the U.S. Office of Education's response to a requirement of the Civil Rights Act of 1964 to investigate the extent of inequality (by race, religion, or national ori-

gin) in the nation's schools. The study's fundamental contribution was to direct attention to the distribution of student performance—the outputs with which I am concerned here. Instead of addressing questions of inequality simply by producing an inventory of differences among schools and teachers according to race and region of the country, the Coleman Report sought to provide an understanding of which, if any, of the observed differences in school resources were important for student learning. This very different perspective—the right one when student educational performance is the concern—set a standard in the school policy debate. Unfortunately, in the subsequent financial equity debate, this important innovation has been largely ignored.

The Coleman Report was widely interpreted as finding that schools are not very important in determining student achievement. Families and, to a lesser extent, peers were seen to be the primary determinants of variations in performance. The findings were clearly controversial and immediately led to a substantial research effort to compile additional evidence about the relationship between school resources and school performance.¹¹

The underlying model guiding the Coleman Report and most subsequent studies is very straightforward. It postulates that the output of the educational process—that is, the achievement of students—is related directly to a series of inputs. Policy makers directly control some of these inputs—for instance, the characteristics of schools, teachers, and curricula. Other factors, such as families and friends plus the innate endowments or learning capacities of the students, generally cannot be affected by public policy. Further, although achievement is usually measured at discrete points in time, the educational process is cumulative; past inputs affect students' current levels of achievement.

Based upon this model, statistical techniques, typically some form of regression analysis, are employed to identify the specific determinants of achievement and to make inferences about the relative importance of the various inputs into student performance. This summary highlights the overall findings from the research.

These studies of educational production relationships measure output not only by student scores on standardized achievement tests but also by other quantitative measures, such as student attitudes, school attendance rates, and college continuation or dropout rates. The general interpretation is that they are all plausible indicators of future suc-

cess in the labor market. This interpretation has been confirmed by other research into labor market outcomes (see review in Hanushek, Rivkin, and Jamison 1992).

Empirical specifications of production function models 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' characteristics (education level, experience, sex, race, and so forth), of the school's organization (class sizes, facilities, administrative expenditure, 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, such as the normal administrative records of schools, that were constructed for other purposes.

Empirical Results for Expenditure Effects

The production function approach has been broadly employed to investigate the impact on school performance of the core factors determining expenditure on education. Instructional expenditure makes up about two-thirds of total school expenditures. Instructional expenditure is in turn determined mostly by teacher salaries and class sizes. Finally, in most U.S. school districts, teacher salaries are directly related to the years of teaching experience and educational level of the teacher. Thus, the basic determinants of instructional expenditure in a district are teacher experience, teacher education, and class size. Most studies, regardless of what other school characteristics might be included, 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 the analyses have such common specifications, the effects of the expenditure parameters can easily be tabulated. Here I present data from a reasonably exhaustive search that uncovered 187 separate "qualified studies" found in thirty-eight separate articles or books through the middle of 1988.¹² These studies, while restricted to public schools, cover all regions of the United States, different grade levels,

different measures of performance, and different analytical and statistical approaches. About one-third draw their data from a single school district, while the remaining two-thirds compare school performance across multiple districts. A majority of the studies (104) use individual students as the unit of analysis, whereas the remainder rely upon aggregate school-, district-, or state-level data. The studies are split about evenly between primary schooling (grades one through six) and secondary schooling (grades seven through twelve). Over 70 percent of the studies measure school performance by some kind of standardized test. However, those using nontest measures (such as dropout rates, college continuation, attitudes, or performance after school) are for obvious reasons concentrated in studies of secondary schooling. There is no indication that differences in sample and study design lead to differences in conclusions.¹³

According both to conventional wisdom and to generally observed school policies, each factor should have a positive effect on student achievement. More education and more experience on the part of the teacher cost more and are presumed to improve individual student learning; smaller classes (more teachers per student) are also expected to be beneficial.¹⁴ More spending in general, higher teacher salaries, better facilities, and better administration should also lead to better student performance. The quantitative magnitudes of estimated relationships are ignored at this point, and attention is focused on the direction of any estimated effect.

The data in table 1 provide a picture of how well conventional wisdom and common school policies hold up to analysis. The columns in the table divide the available estimates by direction of effect and statistical significance. Since not all studies contain estimates of each expenditure component, the first column simply indicates the total number of estimates available. Thus, for example, 152 of the 187 studies include an estimate of the effect of teacher-pupil ratios, or class sizes. Of the 152 estimates of the effects of class size, only 27 are statistically significant. Of these, only 14 show a statistically significant positive relationship, whereas 13 display a negative relationship.¹⁵ An additional 125 estimates show that class size is not significant at the 5 percent level.¹⁶ Nor does ignoring statistical significance help to confirm the benefits of small classes, since the insignificant coefficients have the "wrong" sign by a 46 to 34 margin.¹⁷

The entries for teacher education tell a similar story. The statistically significant results are split between positive and negative relationships, and in a vast majority of cases (100 out of 113) the estimated coefficients are statistically insignificant. Forgetting about statistical significance and looking just at estimated signs again does not make a case for the importance of added schooling for teachers.¹⁸

Table 1. Summary of the Estimated Relationship Between Student Performance and Various Components of School Expenditure (187 studies)

Input	Number of studies	Statistically significant		Statistically insignificant			
		+	-	Total	+	-	Unknown
Teacher/pupil	152	14	13	125	34	46	45
Teacher education	113	8	5	100	31	32	37
Teacher experience	140	40	10	90	44	31	15
Teacher salary	69	11	4	54	16	14	24
Expenditure/pupil	65	13	3	49	25	13	11
Administrative input	61	7	1	53	14	15	24
Facilities	74	7	5	62	17	14	31

SOURCE: Hanushek (1989).

Teacher experience is slightly different. A clear majority of estimated coefficients point in the expected direction, and about 29 percent of the estimated coefficients are both statistically significant and of the conventionally expected sign. But these results only appear strong relative to the other school inputs; they are hardly overwhelming in an absolute sense. Moreover, they are subject to interpretive questions. Specifically, these positive correlations may result from senior teachers having the ability to locate themselves in schools and classrooms with good students. In other words, causation may run from achievement to experience and not the other way around.

Overall, the results are startlingly consistent. No compelling evidence emerges that teacher-pupil ratios, teacher education, or teacher experience have the expected positive effects on student achievement. There is no reason for confidence that hiring teachers with more education or having smaller classes will improve student performance. Evidence of the effect of teacher experience appears marginally more convincing, at least when no consideration is given to the magnitude of any relationship.

The remaining rows of table 1, summarizing information on other expenditure components, including administration, facilities, teacher salaries, and total expenditure per student,¹⁹ provide poorer evidence on the relationship of resources and performance, but what evidence does exist is consistent with the previous results. The quality of administration is measured in a wide variety of ways, ranging from characteristics of the principal to noninstructional expenditure per pupil. Similarly, the quality of facilities is identified through spending and many specific physical characteristics. If only because of the preponderance of positive signs among the significant coefficients, administration appears marginally stronger in its relationship to student achievement than facilities. Nevertheless, the available evidence on both again fails to support convincingly the conventional wisdom.

Finally, and not surprisingly, explicit measures of teacher salaries and expenditure per student do not indicate that they play an important role in determining achievement.²⁰ After all, the underlying components of these expenditures were themselves unrelated to achievement. While negative expenditure effects—in which funds are not only unproductive but also harmful—are difficult to interpret, it is much easier to believe that differences in spending have little or no impact on student performance.

Without systematic tabulation of the results of the various studies, it would be easy to conclude that the findings are inconsistent. But there is a consistency, though it does not match the conventional wisdom. *The research reveals no strong or systematic relationship between school expenditure and student performance.* This is the case both when expenditures are decomposed into their underlying determinants and when they are considered in the aggregate.²¹

Given the general biases toward the publication of statistically significant estimates, the paucity of results confirming the conventional

wisdom is notable. The common calculation of statistical significance is inappropriate when a series of sequential tests of alternative formulations of the achievement relationship is conducted. A sequential approach built on the calculated statistical tests will yield biased estimates of significance. In reality, too many estimated parameters will be judged to be significant.²²

These results reflect the structure and operating procedures of schools observed in existing settings. A different organizational structure with different incentives could produce very different results. For example, almost every economist would support the position that increasing teacher salaries would expand and improve the pool of potential teachers. However, whether this improves the quality of teaching depends on whether or not schools can systematically choose and retain the best teachers from the pool. The results on salary differentials presented previously might be very different if schools faced a greater incentive to produce student achievement and if mechanisms for teacher selection were altered. In other words, there seems little question that money *could* count. It just does not systematically do so with the current organization of schools.

Moreover, the consistency criterion used to judge the results and the potential for policy improvements does not suggest that money never counts. The results are entirely compatible with the notion that some schools use funds effectively and others do not. But unless some way is found to change the districts that would squander additional funds into districts that would use them effectively, added resources are not likely to lead to any improvement in average performance. Good uses of funds are balanced by bad uses within the current structure.

Other Inputs into Education

Since the publication of the Coleman Report, intense debate has surrounded the fundamental question of whether schools and teachers are at all important to the educational performance of students. The Coleman Report has been commonly interpreted as finding that variations in school resources explain only a negligible portion of the variation in student achievement. If this were true, it would not matter which particular teacher a student had or which school a student attended—a conclusion that most people would have difficulty accepting.

The findings of direct analyses of differences among teachers are unequivocal and indicate a very different conclusion: teachers and schools differ dramatically in their effectiveness. A number of studies provide analyses of the differential effectiveness of teachers and schools based on estimation of the average gain in performance of each teacher's (or school's) students.²³ These studies confirm that there are striking differences in teacher performance as measured by average gain in student achievement.

The faulty impressions about the nonimportance of teachers and schools left by the Coleman Report and a number of subsequent studies are the result of a confusion between measures of effectiveness and true effectiveness itself. In other words, existing measures of the characteristics of teachers and schools are seriously flawed and thus are poor indicators of true effectiveness; when these measurement errors are avoided, schools are seen to have important effects on student performance.

These input-output analyses have also investigated a wide variety of other school and nonschool factors. Although it is difficult to be specific in any summary of other factors because the specifications of the various inputs employed in the statistical analyses vary widely, three generalizations are possible. First, family background is clearly very important in explaining differences in achievement. Second, while considerable attention has been given to the characteristics of peers or other students within schools, the findings about their effects are ambiguous. Finally, studies have examined many additional measures of the effects of schools, teachers, curricula, and especially instructional methods on achievement, but no simple characterization of good teachers emerges.²⁴

While not systematically addressed by existing research, one plausible interpretation of the combined results of these studies is that an important element of "skill" is involved in being a successful teacher.²⁵ Skill refers simply to the ability of some teachers to promote higher achievement among their students. The evidence previously presented then indicates that it is currently impossible to identify, much less to measure, components or elements of this skill with any precision. Moreover, the direct evidence casts doubt on whether any form of teacher training course could be organized to foster high skill levels in

teachers. In simplest terms, if we cannot define or measure it, how can we teach it?

Implications for Equity and School Finance Reform

I now turn to the application of this evidence to consideration of school finance reform. Here I sketch some obvious and some less obvious implications of the preceding findings. Again, while school finance policy frequently contains many state-specific nuances, this discussion concentrates on two common elements of “reform.” Most school finance reform programs, based on simplistic equity notions, assume that a basic objective is to limit local variations in school expenditure or, if variations are to exist, to insure that such variations are not related to the property wealth of the district.

The Central Implication

The evidence on school performance indicates that variations in school expenditure are exceedingly poor measures of the variations in education provided to students. Most directly, when students’ learning is the concern, the conventional evidence about inter-district disparities in spending does not identify where educational deficiencies are to be found, and such evidence is generally irrelevant for either an equal protection or an educational disparity court case.²⁶ Such evidence about expenditure simply does not indicate differential provision of education. Therefore, showing how expenditures vary, either absolutely or in accordance with characteristics of districts and students, does not have much use.

We must be quite precise about the interpretation of expenditure. As previously noted, most economists, including myself, would readily accept that differences in spending would be directly related to the education provided *if schools were operating efficiently*. The previously presented evidence indicates clearly, however, that assuming efficiency in spending is entirely inappropriate.

While there are many alternative ways to define and measure educational equity, only the most narrow of these would call for paying

attention to expenditure variations in the face of the evidence that such expenditure variations are unrelated to the education provided. The standard employed would have to be a rigid one linked to dollars, with total disregard for the quality of schooling received by students.

In other words, equity and efficiency are inextricably linked. It is not possible to ignore efficiency issues under the guise of being concerned solely with equity.

Other Implications

There is another side to this discussion: What is likely to happen if we disregard the evidence on the interpretation of expenditure differences and simply make policy on the basis of expenditure differences? This consideration is prompted by a few arguments that are sometimes heard, such as: "The educational problem of the poor is serious, and equalizing expenditure cannot hurt;" or "We should at least give everyone the same chance to make mistakes." The policies flowing from such notions do, unfortunately, have a down side to them.

First, a likely reaction to any move to lessen variation in expenditure is to increase the total level of expenditure on schools. The reason is simple: a state legislature, faced with a need to alter expenditure patterns, finds it much easier to redistribute a larger pie than a fixed pie. In the school finance debate, this is frequently referred to as "leveling up," or bringing the low-spending districts up to the spending levels of the top districts. The arguments behind the policy are generally based either on the need to do better or on pure political necessity. On the other hand, because of the potential for disruption and the obvious divisiveness of "leveling down," there is seldom much interest in this idea. The previous evidence indicates, however, that added funds will on average be dissipated on things that do not improve student achievement (at least unless other, larger changes are also made). Teachers, administrators, and perhaps taxpayers in some districts gaining funds will probably be happier, but the average state taxpayer and parent will not find that the resulting changes do much more than increase tax bills.

Second, there is no assurance that the new funds will go to the schools of poor children. As indicated previously, one of the pervasive views of finance "reform" is that poor children will be helped (or at

least will have a better chance by virtue of greater funding). However, reform schemes designed to follow district wealth patterns can lead to unexpected outcomes because frequently there is not a strong relationship between district wealth and the concentration of student poverty. Some states find that wealthier districts in terms of property wealth per student also have concentrations of poorer families and children. New York State provides a good illustration. Consider the six largest cities in New York State: New York City, Buffalo, Rochester, Yonkers, Syracuse, and Albany. Albany and Yonkers have tax bases in which real property per student is greater than the state average; New York City, Rochester, and Syracuse have tax bases per student only slightly below the state average; and Buffalo is left with a tax base 30 percent below the state average. Yet all of these districts except Yonkers have poverty rates for children above the state average. For example, while the average poverty rate in New York State for children 18 or younger in 1980 was 19 percent, it was over 36 percent in New York City and over 30 percent in Buffalo.²⁷ The largest districts in the state intervened (unsuccessfully) on the side of the plaintiffs in the *Levittown* case and introduced a new argument, municipal overburden,²⁸ in order to protect their funding. In other states, property wealth and poverty may be negatively correlated—that is, high property wealth tends to be found in districts with a small poverty population, but even in these states the overall pattern clearly does not hold jurisdiction by jurisdiction.²⁹ Therefore, while not inevitable, it is likely that many districts serving poor children are hurt in spending terms by plans to neutralize expenditure on the basis of district wealth. Moreover, because of a combination of federal and state grants, districts with concentrations of poor students frequently have above average spending, regardless of their property wealth or overall economic health.³⁰ Programs to limit variations in expenditure could operate to cut back existing compensatory spending for disadvantaged students.

Third, spending differences may not even accurately reflect the real resources each district is able to deliver (i.e., the actual educational inputs). This is the simple result of possible cost differentials facing individual districts. That is, if districts face different prices for things they might buy, from teachers to buildings and equipment, dollar variations themselves do not indicate variations in available real resources. As a simple example, if the schools in one city were less pleasant and

desirable than those in other cities, it would be necessary to pay a higher salary to hire a teacher of equal quality.³¹ An extension of this notion involves districts faced with concentrations of students who are more difficult to educate because of a variety of pre-existing educational deficiencies. These, like cost differences for inputs, lead to expenditure variations in districts behaving in an otherwise identical manner. (Indeed, many state funding formulae recognize such issues and attempt to adjust for input cost differences or for differences in student preparation, handicap status, and the like, even though the magnitude of any real cost differences is poorly understood).

Fourth, districts themselves are not entities to which educational policies should be geared. Individuals choose among districts when they enter an area and move among districts after they live in an area. In fact, there is extensive evidence that individuals make choices among districts in part to satisfy their demands for various public services. Some people who place considerable weight on schooling search for districts that seem to emphasize quality schooling. Others who emphasize other goods or even low public expenditure seek districts that provide an agreeable level and pattern of the services they are looking for. Certainly this system has some drawbacks. Moving can be expensive, and some might find it difficult to move to the districts they would like, for example, because of housing prices, commuting costs, or discrimination. Nevertheless, the fundamental fact for this discussion is that individuals generally have considerable latitude in choosing schools. They are not inextricably tied to a particular district and are not doomed to whatever expenditure levels currently exist in a specific district. Finally, individual districts change their expenditure in line with the desires of the population and with population shifts, so that districts may increase or decrease their expenditure over time. For example, it is possible to trace the movement of district expenditure in the State of Indiana between 1977 and 1987. Only forty-three of the seventy-six top spending districts in 1977 remained in the top quartile in 1987; only forty-two of the seventy-six bottom quartile districts remained there from 1977 through 1987. Thus, policy discussions that speak generally of the population as captives of districts with undesirable spending patterns tend to miss an extremely important feature of the political economy of local jurisdictions. (The special problems of "mobility-constrained" groups, such as the poor, are discussed below).

Fifth, the preferences and movements of citizens across district boundaries have direct ramifications for the observed distribution of property wealth. Specifically, districts that appear to offer a particularly favorable tax and school quality package will appear relatively attractive to many people. This will lead to a bidding up of housing prices in such desirable jurisdictions, because they are in demand, other things being equal. In fact, it is well documented that “otherwise identical” houses will sell for different amounts *because of* citizens’ evaluations of the taxes and the schooling being offered. (See Tiebout 1956, Oates 1969, Rosen and Fullerton 1977, and Wendling 1981). Another way of saying this is that some people pay for their schooling up front through the capitalization of school advantages into the price of homes. Some places that initially look attractive from the vantage point of the tax rate alone are really less attractive because the low rate is multiplied times a high valuation (relative to the other attributes of the home). This has, among other things, a direct effect on the property tax base of the community—something that is often entered into the discussion of the “inequities” of the school finance system. Moreover, reform changes in the funding formula of the state imply distributing somewhat arbitrary capital gains and losses across the jurisdictions in the state. Some places will be made more fiscally attractive and some less by major changes in the financing laws, leading to changes in the capitalization of fiscal differences.

Sixth, in most states spending levels reflect a wide variety of things, including the preferences of the citizens. While it is common to argue that local property wealth is the primary determinant of expenditure differences, that simply is not the case. For example, even though New Jersey and Indiana have relied on local property taxes to fund schools, rough estimates indicate that less than a fifth of the variation in expenditure would be eliminated by totally equating local property wealth per student.³² A combination of local preferences, differences in student needs, curricular choices, cost differentials, and a variety of other factors completely dominate property wealth in the determination of the pattern of expenditure.

Seventh, differences in tax rates across communities bear no direct relationship to the degree of educational equity. Most importantly, school finance reform has been based on perceived differences in the quality of education available, and the quality of education is not

related in any simple way to tax rates. The tax rate provides an indication of the price that residents face to raise funds for schools, and high tax rates might indicate that some districts find it more difficult than others to raise funds through the property tax. But tax rates differ according to a variety of factors, including community preferences, community income and wealth, the amount of nonresidential wealth in the tax base, and so forth. The pattern of tax rates may be an issue from the standpoint of various notions of "taxpayer equity," but tax rates seldom have much to do with considerations of equity in education. Further, while the education clauses of state constitutions may require states to provide certain levels of education, they never indicate that school tax rates must be equalized across a state.

This list of likely ramifications underscores the point that simple alterations in expenditure patterns can have consequential and undesirable effects. What is already known about the educational process and about behavior of local jurisdictions leads to the inescapable conclusion suggested in my introduction: the general assumptions behind early school finance reform are misleading at best.

Magnitude of Expenditure Effects

The evidence presented in Table 1 did indicate that a majority of studies found a positive relationship between aggregate expenditure and student performance, albeit few statistically significant relationships. While this finding might suggest a potential for equity improvement by means of adding resources to low-spending districts, moving to such policy deliberations requires consideration of the magnitude of any expenditure effects. Specifically, how much could achievement in low-spending districts be altered by an infusion of new resources?

Two alternative estimates, representing very different circumstances, illustrate why the magnitude of performance change associated with expenditure increases must enter into policy considerations. First, Wendling and Cohen (1981) conducted a study of expenditure effects in 1977-78 in New York State, the state with the second highest rate of spending (behind Alaska) during that year.³³ Second, new evidence for 1991 from the State of Alabama, with the 46th highest spending rate in 1990, provides information about expenditure relationships at low levels of expenditure.

New York has consistently been at or near the top of spending on schools across all states. Its average expenditure in 1978 was 45 percent above the average for the country. Wendling and Cohen (1981) examine whether or not expenditure differences among districts in New York matter for student achievement, and they conclude that indeed expenditure is important. They analyze average third-grade reading and mathematics achievement for 1,021 districts. While they examine various model formulations and different groupings of districts,³⁴ the results for the entire state are representative and provide a clear indication of how expenditure relates to student performance.

The estimated effect of approved operating expenditure³⁵ on student performance is uniformly statistically significant, but the magnitude of the estimates shows the difference between statistical significance and policy significance. The estimated expenditure parameters are .001 and .002 for third-grade reading and mathematics performance, respectively. This implies that a \$1,000 increase in expenditure per student yields an additional point on the reading test and an additional two points on the mathematics test. While absolute scores are difficult to interpret, some idea of magnitude can be gained by looking at movement in the distribution of spending and performance. An increase of \$1,000 is a 50 percent increase in state school spending and is 2.2 standard deviations in expenditure within the sample of schools, but it yields an increase in performance of only .2 to .3 standard deviations. This is equivalent to moving the average student up to around the sixtieth percentile or to moving a student starting at the 10th percentile to the 15th percentile. In simple terms, attempting to increase performance through simple increases in expenditure is very expensive.

Alabama is a relatively low-spending state, falling at the other end of the distribution from New York State. The state's highest-spending district in 1991 (Mountain Brook) had total current expenditure per student of \$5,113. This is slightly below the mean level for the entire nation in 1991, for which the preliminary estimates are \$5,237. The minimum spending in Alabama was about \$2,900 per student. Thus, variations in spending in Alabama should provide some insight into whether or not there exists some threshold expenditure below which changes in spending have strong and noticeable effects on achievement—the issue that comes up in discussions of “adequacy” (Celis 1992).

The statistical analysis attempts to explain variations across districts in Alabama's Basic Competency Tests (BCTs). These are criterion-referenced tests adopted by the state board of education to measure whether or not students are accomplishing what was expected of them according to the curriculum for different grade and subject areas (reading, mathematics, and language arts). The performance measure is the percentage of a district's students meeting the minimum standards for the specific tests ("passing").³⁶ Weighted least squares regression analysis is employed to estimate the effect of current expenditure per student in average daily attendance (ADA) on performance, while holding constant the influence of family background and school district type (i.e., city or county district).³⁷

The results of estimates for the State of Alabama can be easily summarized. Table 2 presents the estimated expenditure effects. None of the nine estimated relationships is statistically significant at the 5 percent level, and one is even negative.³⁸

Table 2. Estimated Change in BCT Pass Rate for Expenditure Increase of \$1,000 per Student in Average Daily Attendance (ADA): Alabama, 1990-91

	Test and Grade								
	Read gr 3	Read gr 6	Read gr 9	Math gr 3	Math gr 6	Math gr 9	Lang gr 3	Lang gr 6	Lang gr 9
Change in percent passing	0.568	0.119	2.733	2.562	0.423	1.574	1.656	-0.133	2.565
t-ratio	0.51	0.07	1.95	1.70	0.19	0.78	1.00	-0.07	1.54

NOTE: Estimates are the weighted least squares regression estimates of the determinants of percent passing the Alabama Basic Competency Test (BCT) for different subject areas and grades for 127 school districts in 1990-91 (Hoover and Mt. Brook excluded). Estimates equations include percent of students receiving free or reduced lunch; percent nonwhite; city school district indicator; and current expenditure per ADA. Weights are the number of students in average daily attendance.

Abbreviations:

ADA: per student in Average Daily Attendance

BCT: Basic Competency Test

Table 3 presents the results of two alternative policy scenarios. The first brings all of the spending of districts below the median in the state up to the median. The second, which is sometimes called "full leveling

Table 3. Predicted Change in State Pass Rates from Increased Spending: Alabama Districts, 1990–91

A. Partial leveling-up by bringing all low-spending districts to median (cost=\$74 million)									
	Reading			Mathematics			Language		
	grade 3	grade 6	grade 9	grade 3	grade 6	grade 9	grade 3	grade 6	grade 9
1990–91 pass rate	81.3	64.3	73.1	78.8	59.4	43.8	73.2	59.7	52.2
Change in pass rate	0.1	0.0	0.3	0.3	0.0	0.2	0.2	-0.0	0.3
Projected pass rate	81.4	64.3	73.4	79.1	59.4	44.0	73.4	59.7	52.5
B. Full leveling-up by bringing all districts to top (cost=\$1.05 billion)									
1990–91 pass rate	81.3	64.3	73.1	78.8	59.4	43.8	73.2	59.7	52.2
Change in pass rate	0.9	0.2	4.2	4.0	0.7	2.4	2.6	-0.2	4.0
Projected pass rate	82.2	64.5	77.3	82.8	60.1	46.2	75.8	59.5	56.2

up,” brings all districts up to the level of the highest spending district in the state. Bringing all states to the median expenditure level would cost an additional \$74 million (compared to total spending of \$2.4 billion). While this increase of 3 percent is not a large relative change in expenditure, the top panel of table 3 indicates that it would yield imperceptible changes in performance on most of the tests. The bottom panel provides estimates of the achievement effects of full leveling up—i.e., bringing all district spending up to that of Mountain Brook. Such a policy would cost \$1.05 billion and would yield at most a 4 percentage point increase in students passing the BCT in the state. The net impact of leveling up is shown in figure 2.

Figure 2
Alabama Performance with Leveling-Up

The importance of these results is clear. A policy of bringing all districts to the top in spending would place Alabama schools at approximately the mean for the nation, up dramatically from its current

position in the national distribution of spending. But this increase of the state's school spending by over one-third would have very small effects on aggregate school performance, at least if the schools behave in a way consistent with current practice. The resulting performance would remain very far below the state's goals of a 95 percent passing rate on the separate BCT tests.

There is also no evidence from analyzing the schools of Alabama that there exists a minimum threshold for school spending. There is no apparent range of stronger influence of spending on achievement than is found for the entire set of schools.

Policy Alternatives

Concerns about the implications of school finance reform do not, of course, vitiate the undeniable need to improve our public schools. The deplorable conditions described in Kozol (1991) require addressing. The intentions of finance reformers have been, in my opinion, good. Only their approach is questionable. Three general factors lead to the judgment that structural change is essential. First, in absolute terms students are not performing up to expectations. Performance, as measured by standardized tests over time, international comparisons of tests, various measures of workplace performance, and common perceptions, is currently unacceptable. Second, as indicated by the previous evidence, there is overwhelming evidence that the resources devoted to schools—which have been both large and increasing—have not been effectively used. Third, the significantly skewed distribution of educational success, which leaves poor and minority students behind the rest of the population, is incompatible with most people's views on the goals of our society.

The previous sections of this paper indicate why "reform" as commonly included in school finance considerations is unlikely to address any of these causes of concern. The primary focus on the distribution of state financial aid or the limits on local fiscal options distracts attention from the issues of school organization, incentives for performance, and the goals of the system. Because of the contentiousness of issues surrounding the distribution of funds, school finance debates

have the potential for absorbing all energy related to school policy. Thus, in addition to generally offering few solutions to the problems previously identified, there is a significant opportunity cost in stalling some of the fundamental restructuring that must proceed if we are to deal with the current problems of schools. This problem of distraction, of course, is not inevitable, but there are strong forces pushing in that direction.

The concentration of courts and legislatures on finance reform while skirting away from more fundamental policy considerations does follow a certain logic. Expenditures are readily measurable; there is a plausible argument behind their importance; there is no obvious alternative focus of policy; and operating on expenditure at least represents doing something. In other words, there are serious problems, and it is perceived that at least some attempt to remedy them should be undertaken.

This logic is supported by the lack of convincing evidence that any specific approaches or policies will bring about significant improvements in student performance. As reviewed previously, no set of simple changes involving either resources or programs shows a consistently strong relationship with performance.

But that is just the problem. Concentrating solely on dollars or resources does not confront the basic structural problems in the operations of schools. For example, while Jonathan Kozol (1991) points to the lack of achievement of impoverished students and calls for increased funding, he ignores the fact that the increased funding will be administered by the same school boards and administrators that he railed against in *Death at an Early Age* (Kozol 1967). An alternative approach, which suffers from many of the same problems, is to argue that additional funds would not be utilized in old unproductive ways but would be used in highly directed ways that insured achievement. This approach is often accompanied by the description of a specific program that has been shown to work in the few places where it has been tried. The problem, of course, is that we really do not know in general terms what will work, and the successful programs that have been identified have not been broadly introduced by districts with the funds to do so (for example, through general federal compensatory education funding or more generous local support). There is simply no reason to believe that a centrally directed system of increased funding

relying on the simple identification of productive programs or spending patterns would be more productive than what has been currently observed.

An alternative formulation of educational policies avoids the pitfalls of previous approaches and offers considerably more promise of improvement. The alternative is moving to organizations and incentive systems that directly reward performance. The current set of policies, almost exclusively pursued, involves providing or requiring certain inputs—expenditures, class sizes, teacher attributes, and the like. These input policies are essentially pursued and continued without regard to their effectiveness, either in the aggregate or in specific instances. The proposed alternative is to concentrate on student performance instead of factors thought (or hoped) to be important in determining student performance.

Various systems have been used or suggested to promote performance-based policies, including merit pay for teachers, merit awards for schools that perform well, and a variety of plans emphasizing choices among educational institutions. Essentially, the common ingredient of such plans is that resources are directly related to performance: if performance is high, resources are high; if performance is low, resources are commensurately low. For example, merit pay for teachers operates by increasing salaries of those who perform well and not of those who perform poorly. Similarly, choice plans, which operate by allowing students and parents to choose among alternative schools, work by reinforcing parental judgments about quality schools, with suitable flows of resources to the schools that attract students.

The orientation of these policies is based on finding the correct incentives. If tangible incentives for improved performance are offered, most decision making can be expected to improve. Actual operations of hiring, promotion, curriculum, student placement, and the like—while not specified or regulated by a central authority—can be expected to respond to incentives. This has been demonstrated by wide-ranging research, both in education and elsewhere. Determining effective incentives, however, will take experimentation, bargaining, and evaluation.

There are many different versions of these performance-based plans, particularly of the choice plans. Commonly discussed choice plans range from magnet or special schools (which are fairly wide-

spread in some urban districts) to full vouchers, which provide parents with funds to pay tuition at either public or private schools of their choice. Between these extremes are free choice within public school districts, open enrollment in any public school in the state, and tuition tax credits to rebate a portion of any tuition payments to the parents.

Performance-based options have been discussed widely and will not be reviewed here (see Chubb and Hanushek 1990). Instead I will merely highlight two features. Each option has conceptually appealing elements. And there is little historical evidence for each option that would provide details of either how it should be implemented or the magnitude of gains that might be expected. In other words, there is also considerable uncertainty, particularly about details of implementation, because these approaches are largely untried. The uncertainty should not, however, be taken as a reason for avoiding them but should dictate a more interactive approach to policy making instead. Moreover, as indicated, each performance-based option has considerable appeal, especially as an option to the almost universally employed input policies that have had such a dismal record.

The performance-based view of educational policy is very different from the current view of how to make policy. It also is not very amenable to the simple remedies and simple tracking of responses so appealing within a court context. Nevertheless, for all its messiness and uncertainty, it offers some realistic hopes for improvement—something that is absent from narrow decisions on expenditure and other inputs.

The use of performance-based plans is supported by the research into educational performance reviewed above. This research indicated extremely large and significant variations in the performance of individual teachers and schools. It is also very important to reiterate here that research has concentrated on the value-added of teachers and not on absolute performance levels of students. The research demonstrates that there can be low value-added in a “good” suburban school where the absolute level of achievement is quite high. Similarly, there can be high value-added within “bad” central city schools where students come to school quite unprepared but leave with marked increases in their achievement (see, for example, Hanushek 1992). While research cannot identify the components of successful and unsuccessful teaching, it supports the simple but powerful notion that good performance

can be identified by school administrators (see Murnane 1975 and Armor et al. 1976). Further, if this capacity can be extended to individual parents—who frequently at least act as if they can tell the difference between good and bad teachers—the groundwork for performance-based policies is established.

The overall point is straightforward: a range of effective policies appears to be available. They are, however, almost certainly very different from the traditional policy focus and the orientation of traditional school finance “reform” efforts. Moreover, instituting some of these fundamental reforms might take added funds, particularly in the implementation and learning phases. There is a huge difference, as should be clear, between expenditures directly linked to improved incentives and student performance and expenditures made in the hopes that something good will happen.

Finally, the restructuring of incentives in schools appears to be the only feasible answer to dealing with the gloomy record schools have in improving the performance of educationally and economically disadvantaged youth. Various input-oriented programs have been mounted to deal with the disadvantaged, including a large portion of all federal spending on schools, but there is little evidence that this has had much impact. The alternative to restructuring incentives as proposed here is to continue to expand the programs that have thus far been unsuccessful.

The evidence from past analyses demonstrates that good teachers exist in what are commonly thought to be bad urban districts. Their existence, however, is masked by generally low achievement levels; that is, even though an individual teacher may elicit more than one year of achievement growth within a one-year period, low absolute levels of performance could hide it. The policy problem is that we have not been able to attract, to identify, and retain sufficiently large numbers of such good teachers so as to have the kind of influence that is needed. This is just the appeal of performance-based incentive schemes. They are designed to reinforce good performance. We should, at the same time, not have overly optimistic expectations. As has been thoroughly documented, family influences are very powerful in determining achievement levels; so while specific teachers might have a substantial influence on achievement, they might not immediately overcome the deficits arising from factors outside the schools.

Indeed, it may take the continued efforts of many good teachers over the course of the student's school career. This reality, however, should not deter our efforts to provide the best possible education.

Conclusions

School finance reform, as commonly espoused in courts and legislatures across the country, is likely to work against the very improvements most needed in public schools. By its nature, emphasizing primarily the distribution of expenditure per student, financing reform is almost certain to exacerbate existing problems of inefficiency in school operation.

Discussions of school finance reform typically attempt to separate considerations of efficiency from issues of "equity." Such a distinction is impossible, however, if the definition of equity involves the learning of children, which depends directly on the ability of school districts to translate resources into student achievement. If schools are ineffective at this task, little can be done to improve equity in student performance by simply heaping more resources on poorly performing districts.

Research into the relationship between resources and student performance, conducted over the past quarter century, has demonstrated conclusively that, *within the current organization and operation of schools*, there is no consistent relationship between resources and student performance. Common policy proposals—ones that are used to justify pleas for added resources to school districts—simply are not supported by evidence about their performance within schools.

Ignoring the evidence on performance is likely to worsen the problems of performance and inefficiency. The current incentive structure in schools does not promote efficient use of resources. Therefore, while additional funds might be used effectively by some districts, other districts will probably use them ineffectively—which, if past history is a guide, leads to no aggregate improvement in performance from increased funds.

Large differences in performance exist among teachers and schools. Past evidence further indicates that parents and administrators can identify the best and worst among these. What is missing is an effec-

tive structure for channeling knowledge about performance into overall improvement in the schools. A variety of mechanisms for this have been proposed; although there is little operational experience with them, the key to each is that resources are more directly linked to performance of the schools. This is an entirely different perspective from what has been found in most discussions of school finance reform, which gives no weight to student performance in making resource allocations.

NOTES

1. See, for example, Coons, Clune, and Sugarman (1970) or Wise (1968).
2. The term "efficient" here is used in the economist's sense of obtaining the maximum possible performance from any given expenditure of resources. This definition is very different from that which appears to have been employed in a number of legal arguments emanating from state constitutional requirements to provide an efficient system of public schools.
3. See, for example, Coons, Clune, and Sugarman (1970, p. 29) where they discuss Ribich (1968). They state, "Ribich's painstaking analyses suggest, if anything, a variety of sometimes conflicting relationships between cost and purely economic benefits from added dollar increments." They go on to indicate, "There are similar studies suggesting stronger positive consequences from dollar increments, and there are others suggesting only trivial consequences, but the basic lesson to be drawn from the experts at this point is the current inadequacy of social science to delineate with any clarity the relation between cost and quality. We are unwilling to postpone reform while we await the hoped-for refinements in methodology which will settle the issue" (p. 30).
4. For example, after discussing the difficulty of employing alternative measures of real resource differences (such as education levels of teachers), Coons, Clune, and Sugarman (1970, p. 26) state: "We have no stomach for such an imbroglio. Ultimately we will need a standard appropriate to the rigors of judicial proof, and the only convincingly quantifiable item in the spectrum is money available for the general task of education in each district."
5. This section draws extensively on the presentation in Hanushek (1989) which in turn updates previous analyses in Hanushek (1981, 1986).
6. Current expenditures per student are deflated by the consumer price index. See *Digest of Education Statistics*, 1989, tables 88 and 114, and updates.
7. On the other hand, evidence from international tests in 1964 suggest that U.S. students have historically done relatively poorly (Husen 1967).
8. *Digest of Education Statistics*, 1988, table 51.
9. Moreover, only 3 percent of teachers in 1986 were in their first year of teaching (*Digest of Education Statistics*, 1988, table 54). The aging and stagnation of the teaching force have, however, been the subject of separate concerns.
10. The teacher salary data over time that are normally cited are provided by the National Education Association, and the sample and reliability of these are unknown (*Digest of Education Statistics*, 1988, table 57). An alternative source, the decennial population censuses, indicates smooth increases in salaries of teachers by decade, but these fall relative to annual earnings of all college graduates. See Hanushek, Rivkin, and Jamison (1992).

11. There were also extensive analyses of the report's methodology and of the validity of its inferences. See, for example, Bowles and Levin (1968), Cain and Watts (1970), and Hanushek and Kain (1972).

12. A qualified study was defined as a production function estimate that: (1) is published in a book or refereed journal; (2) relates some objective measure of student output to characteristics of the family and the schools attended; and (3) provides information about the statistical significance of estimated relationships. Note that 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 count as only one study). Search procedures for articles involved using the national educational data base (ERIC), searching the bibliographies of included and related articles, and scanning the tables of content of likely journals. No articles uncovered in this search and meeting the above criteria were excluded, but inevitably some were missed. Articles from this time period that were overlooked in the search process but that have subsequently been discovered include Brown (1972), Walberg and Rasher (1974), Wendling and Cohen (1981), and Walberg and Fowler (1987). And, of course, there have been publications subsequent to construction of this table. The inclusion of these articles does not change the weight of the evidence or the substantive conclusions reached.

13. The tabulations, when stratified by grade level, by whether individual or aggregate data were used, by output measure, and by value-added or level forms of estimation, yield the same qualitative conclusions.

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

15. Teacher-pupil ratios are treated here as 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. Since much of the evidence here refers to the situation before such legislation and restrictions, it is reasonable to interpret the evidence as relating to class sizes.

16. In any statistical analysis, which necessarily relies on a sample of all possible students and classroom environments, an estimated relationship may not be real but only perceived to be so because of the specific sample. Standard regression techniques provide ways of estimating the likelihood of being fooled by the sampling into thinking there is a relationship when in fact there is not. The shorthand term, "statistical significance," implies that less than 5 percent of the time when there is really *no* relationship would we get an estimate as large as the one obtained. In other words, when the estimate is "statistically significant," we are quite confident that some relationship does indeed exist. In all cases, however, the estimates of statistical significance assume that the "correct" relationship is being estimated; that is, that the model of achievement is properly specified to include the relevant factors determining performance.

17. Note that not all studies report the sign of insignificant coefficients. For example, 45 studies report insignificant estimated coefficients for teacher-student ratios but do not report any further information.

18. Note that only 113 studies report evidence about teacher education. Since data on teacher education are so readily available, it seems likely that a number of additional studies investigated teacher education effects but discarded the results without reporting them after finding negative or insignificant effects.

19. Information on each of these is less frequently available. This is partially explained by common reliance on administrative records which do not record them (except perhaps teacher sal-

aries). The level of the analysis and sampling frame for some studies offer another explanation; for example, since expenditures per student are generally measured for districts, the analyses that rely on individual student data for a single district would find no variation in this input and thus could not include it. More recent studies have generally concentrated on the analysis of individual student and classroom data and thus have not considered aggregate expenditure effects.

20. The interpretation of expenditure and salary estimates is sometimes clouded by including them in addition to teacher experience, education, and class size. Because multiple regression coefficients indicate the effect of a specific variable when all other variables are held constant, direct measures of expenditure, for example, in models also including the prime determinants of instructional spending would be interpreted as the effect of noninstructional expenditures on achievement. Also, because prices can vary across the samples in the separate studies, it is more difficult to interpret dollar measures than real input measures. Finally, eight of the thirteen significant positive expenditure results in table 1 come from the different estimates of Sebold and Dato (1981). In this study, imprecise measurement of family inputs suggests that school expenditures may in fact mainly be a proxy for family background.

21. There are several obvious reasons for caution in interpreting this evidence. For any individual study, incomplete information, poor quality data, or faulty research could distort the statistical results. Even without such problems, the actions of school administrators could mask any relationship. For example, if the most difficult students to teach were consistently put in smaller classes, any independent effect of class size could be difficult to disentangle from the mismeasurement of a students' characteristics. Finally, the statistical insignificance of estimates can reflect no relationship, but it also can reflect a variety of data problems, including high correlations among the different measured inputs. In other words, as in most research, virtually any of the studies is open to some sort of challenge.

22. This issue is discussed in Hedges (1990). "The published literature is particularly susceptible to the claim that it is unrepresentative of all studies that may have been conducted (the so-called publication bias problem). There is considerable empirical evidence that the published literature contains fewer statistically insignificant results than would be expected from the complete collection of all studies actually conducted. . . . There is also direct evidence that journal editors and reviewers intentionally include statistical significance among their criteria for selecting manuscripts for publications. . . . The tendency of the published literature to overrepresent statistically significant findings leads to biased overestimates of effect magnitudes from published literature. . . ." (p. 19, listed references omitted).

23. These studies are analyses of covariance or, equivalently, regression analysis using individual teacher (or school) dummy variables in addition to measures of prior student achievement, family background factors, and other explicitly identified inputs in a regression format. See Hanushek (1971, 1992); Murnane (1975); Armor et al. (1976); and Murnane and Phillips (1981).

24. Perhaps the closest thing to a consistent conclusion across the studies is that "smarter" teachers, ones who perform well on verbal ability or achievement tests, do better in the classroom. Nonetheless, while plausible, there remains mixed evidence on how good teacher tests are at indicating teaching ability. Tabulations similar to those in table 1 indicate thirty-one studies that have analyzed teacher verbal scores. Of these, eight find positive and significant relationships, and another ten find positive but insignificant relationships. These overall findings have been extended by a recent study by Ronald Ferguson (1991), which finds teacher ability as measured by scores on the Texas teacher test to be related to student performance, although that study is insufficient to change the weight of the evidence.

25. The idea of skill differences among teachers is not the only possible interpretation of the data. Differences in achievement across classrooms could reflect differences in teachers, in other classroom-specific factors, or a combination of both. The teacher skill interpretation is suggested

by the fact that principals' ratings of teachers are correlated with the covariance estimates of classroom differences; see Murnane (1975) and Armor et al. (1976). Evidence on the stability of teacher effects across grades, test area, and years for individual teachers further supports the interpretation based on teacher skill; see Hanushek (1992). A discussion of skill differences in the production function context can be found in Hanushek (1986).

26. School finance court cases have typically contained two elements. First, an equal protection argument is employed, which asserts that school expenditure differences related to variations in the local property tax base are discriminatory. Second, the "education clause" usually found in the state constitution is used to back an assertion that large variations in expenditures are impermissible. In both instances, the direct evidence provided for the alleged wrong involves variations in expenditures (sometimes linked to other things such as property tax wealth).

27. See New York State Office of the Comptroller, *Financial Data for School Districts, 1982* and U.S. Bureau of the Census, *County and City Data Book, 1983* (Washington, DC: U.S. Government Printing Office, 1983), tables A and C.

28. The argument of municipal overburden is that excessive demands for nonschool expenditures faced by urban districts subtract from what otherwise would be available for schools. Therefore, the state funding formula should recognize these other expenditures in allocating school support. See the arguments in *Levittown*. For an economic analysis, see Brazer and McCarty (1987).

29. As described, there is considerable variation in tax bases and poverty rates within a state. Thus, for example, Albany had a property tax base per student that was 34 percent above the state average, and yet it also had a poverty rate above the state average. Cutting back on funds for this "wealthy" district would potentially harm sizable numbers of poor children.

30. For example, in the situation in New York State, each of the six large districts except New York City had expenditures per student above the state average. See *Financial Data for School Districts, 1982*.

31. This situation, known to economists as "compensating differentials," can exist whenever jobs or job locations include different attributes such as riskiness, opportunities for learning, or, in the case of cities, favorable living conditions. For a general description, see Ehrenberg and Smith (1991) or Hamermesh and Rees (1988). In the context of teachers, see Toder (1972), Antos and Rosen (1975), and Kenny and Denslow (1980). Differences in the attractiveness of areas can also lead to differences in housing and land prices, thus affecting other inputs to education. See, for example, Roback (1982).

32. These calculations rely on estimates of the relationship between expenditures per student and wealth per student in districts in these states. The R^2 of simple regression in each state was less than .20, indicating that one-fifth would be an estimate of the upper bound on the potential for equating spending by eliminating property tax base differences.

33. This study is entitled "Education Resources and Student Achievement: Good News for Schools," presumably because it was one of the few studies that ever found statistically significant relationships between expenditure and policy.

34. The basic regression models estimated include median years of schooling and percent below poverty in the district, percent minority students, district size, and pupil/teacher ratio in addition to expenditure. Expenditure is measured in several alternative ways, and some formulations include treatment of geographic location of districts.

35. Approved operating expenditure excludes certain categories included in total current expenditure, such as some transportation, employee benefits, etc. The average approved expenditure in 1978 for the sampled schools was \$2,064, compared to an average total current expenditure for the state of \$2,527.

36. The results are unchanged in a qualitative sense if performance is measured by average scores in the district as opposed to the percentage of students passing the BCT examinations. The magnitude of changes in average scores is somewhat less than of the changes in pass rates predicted for spending changes, a finding that is consistent with the notion that the average performance is relatively close to the established passing score on most of the separate BCTs.

37. Family background is measured according to the percentage of students receiving free or reduced lunch and to what percentage is nonwhite. In 1990–91, there are 129 separate school districts of which 67 serve countywide populations (outside of city districts), while the remaining 62 serve individual cities. The estimates are weighted by the number of students in average daily attendance, in order to deal with the heteroscedasticity introduced by averaging performance across populations of different sizes.

38. The estimates presented exclude two districts (Hoover and Mt. Brook) that are significant outliers in terms of expenditure levels. Because they are noticeably distant from most other districts they have an undue influence on the estimated expenditure effects. Including these districts yields somewhat larger expenditure effects (ranging from 1.0 to 3.7 percent passing per \$1,000 as opposed to -.1 to 2.7 percent passing per \$1,000 in table 2). Three of the nine estimated coefficients are significant at the 5 percent level when the two outliers are included. These estimates do not, however, provide reliable information about the effects of increasing expenditure because there is little or no pattern to expenditures for the remaining 127 districts included in the analysis.

References

- Antos, Joseph R., and Sherwin Rosen. 1975. "Discrimination in the Market for Teachers," *Journal of Econometrics* 2 (May): 123–50.
- Armor, D., P. Conry-Oseguera, M. Cox, N. King, L. McDonnell, A. Pascal, E. Pauly, and G. Zellman. 1976. *Analysis of the School Preferred Reading Program in Selected Los Angeles Minority Schools*. R-2007-LAUDS. Santa Monica, Calif.: Rand Corporation.
- Behrendt, A., J. Eisenach, and W. R. Johnson. 1986. "Selectivity Bias and the Determinants of SAT Scores," *Economics of Education Review* 5, 4: 363–371.
- Bieker, R. F., and K. R. Anschel. 1973. "Estimating Educational Production Functions for Rural High Schools: Some Findings," *American Journal of Agricultural Economics* 55: 515–519.
- Berne, Robert. 1988. "Equity Issues in School Finance," *Journal of Education Finance* 14 (Fall): 159–180.
- Berne, Robert, and Leanna Stiefel. 1984. *The Measurement of Equity in School Finance: Conceptual, Methodological, and Empirical Dimensions*. Baltimore: Johns Hopkins University Press.
- Bishop, John. 1989. "Is the Test Score Decline Responsible for the Productivity Growth Decline?" *American Economic Review* 79, 1 (March): 178–197.
- Bowles, S., and H. M. Levin. 1968. "The Determinants of Scholastic Achievement—An Appraisal of Some Recent Evidence," *Journal of Human Resources* 3, 1: 3–24.
- Brazer, Harvey E., and Therese A. McCarty. 1987. "Interaction Between Demand for Education and for Municipal Services," *National Tax Journal* 40, 4 (December): 555–566.
- Brown, Byron W. 1972. "Achievement, Costs, and the Demand for Public Education," *Western Economic Journal* 10, 2 (June): 198–219.
- Cain, G. G., and H. W. Watts. 1970. "Problems in Making Policy Inferences from the Coleman Report," *American Sociological Review* 35, 2: 328–352.
- Celis, William III. 1992. "Twenty-Three States Face Suits on School Funds," *New York Times*, September 2.
- Chubb, J. E., and E. A. Hanushek. 1990. "Reforming Educational Reform." In *Setting National Priorities: Policy for the Nineties*, edited by Henry J. Aaron. Washington: Brookings Institution, pp. 213–248.
- Cohn, Elchanan. 1979. *The Economics of Education*, Rev. ed. Cambridge, MA: Ballinger.
- Cohn, Elchanan, and Terry G. Geske. 1990. *The Economics of Education*. 3d ed. Oxford: Pergamon Press.

- Coleman, J. S., T. Hoffer, and S. Kilgore. 1982. *High School Achievement: Public, Catholic, and Private Schools Compared*. New York: Basic Books.
- Coleman, James S., Ernest Q. Campbell, Carol J. Hobson, James McPartland; Alexander M. Mood, Frederic D. Weinfield, and Robert L. York. 1966. *Equality of Educational Opportunity*. Washington, DC: Government Printing Office.
- Congressional Budget Office. 1986. *Trends in Educational Achievement*. Washington, D.C.
- _____. 1987. *Educational Achievement: Explanations and Implications of Recent Trends*. Washington, D.C.
- Coons, John E., William H. Clune III, and Stephen D. Sugarman. 1990. *Private Wealth and Public Education*. Cambridge, MA: Belknap Press of Harvard University Press.
- Digest of Education Statistics*. 1989. National Center for Education Statistics. Washington, DC: Government Printing Office.
- Dynarski, M. 1987. "The Scholastic Aptitude Test: Participation and Performance," *Economics of Education Review* 6, 3: 263–274.
- Ehrenberg, Ronald G., and Robert S. Smith. 1991. *Modern Labor Economics: Theory and Public Policy*. 4th ed. New York: Harper Collins.
- Ferguson, Ronald. 1991. "Paying for Public Education: New Evidence on How and Why Money Matters," *Harvard Journal on Legislation* 28, 2 (Summer): 465–498.
- Garms, Walter I., James W. Guthrie, and Lawrence C. Pierce. 1978. *School Finance: The Economics and Politics of Public Education*. Englewood Cliffs, NJ: Prentice-Hall.
- Glass, G. V., and M. L. Smith. 1979. "Meta-Analysis of Research on Class Size and Achievement," *Education Evaluation and Policy Analysis* 1, 1: 2–16.
- Greenberg, D., and J. McCall. 1974. "Teacher Mobility and Allocation," *Journal of Human Resources* 9: 480–502.
- Hamermesh, Daniel S., and Albert Rees. 1988. *The Economics of Work and Pay*. 4th ed. New York: Harper & Row.
- Hanushek, Eric A. 1971. "Teacher Characteristics and Gains in Student Achievement: Estimation Using Micro-Data," *American Economic Review* 61, 2 (May): 280–288.
- _____. 1979. "Conceptual and Empirical Issues in the Estimation of Educational Production Functions," *Journal of Human Resources* 14, 3 (Summer): 351–88.
- _____. 1981. "Throwing Money at Schools," *Journal of Policy Analysis and Management* 1, 1: 19–41.

- . 1986. "The Economics of Schooling: Production and Efficiency in Public Schools," *Journal of Economic Literature* 24, 3 (September): 1141–1177.
- . 1989. "The Impact of Differential Expenditures on School Performance," *Educational Researcher* 18, 4 (May): 45–51.
- . "When School Finance 'Reform' May Not Be Good Policy," *Harvard Journal on Legislation*, 28, 2 (Summer): 423–456.
- . "The Trade-off Between Child Quantity and Quality," *Journal of Political Economy* 100, 1 (February): 84–117.
- Hanushek, Eric A., and John E. Jackson. 1977. *Statistical Methods for Social Scientists*. New York: Academic Press.
- Hanushek, Eric A., and J. F. Kain. 1972. "On the Value of 'Equality of Educational Opportunity' as a Guide to Public Policy." In *On Equality of Educational Opportunity*. Edited by F. Mosteller and D. P. Moynihan. New York: Random House.
- Hanushek, Eric A., Steven G. Rivkin, and Dean T. Jamison. 1992. "Improving Educational Outcomes While Controlling Costs," *Carnegie-Rochester Conference Series on Public Policy*. Vol. 37. Elsevier, pp. 205–38.
- Hedges, Larry V. 1990. "Directions for Future Methodology." In *The Future of Meta-Analysis*. Edited by Kenneth W. Wachter and Miron L. Straf. New York: Russell Sage. Pp. 11–26.
- Husen, Torsten, ed. 1967. *International Study of Achievement in Mathematics*. Vol. 2. New York: John Wiley and Sons.
- Kearney, C. Philip, and Li-Ju Chen. 1989. "Measuring Equity in Michigan School Finance: A Further Look," *Journal of Education Finance* 14 (Winter): 319–67.
- Kenny, Lawrence W., and David A. Denslow, Jr. 1980. "Compensating Differentials in Teachers' Salaries," *Journal of Urban Economics* 7 (March): 198–207.
- Kozol, Jonathan. 1967. *Death at an Early Age: The Destruction of the Hearts and Minds of Negro Children in the Boston Public Schools*. New York: Houghton Mifflin.
- . 1991. *Savage Inequalities: Children in America's Schools*. New York: Crown Publishers.
- McKnight, Curtis C., et al. 1987. *The Underachieving Curriculum: Assessing U.S. School Mathematics from an International Perspective*. Champaign, IL: Stipes.
- McMahon, Walter W., and Carroll Melton. 1978. "Measuring Cost of Living Variation," *Industrial Relations* 17, (October): 324–332.
- Murnane, Richard J. 1975. *Impact of School Resources on the Learning of Inner City Children*. Cambridge, MA: Ballinger.

- . 1981a. "Teacher Mobility Revisited," *Journal of Human Resources* 16, 1 (Winter): 3–19.
- . 1981b. "Interpreting the Evidence on School Effectiveness," *Teachers College Record* 83, 1: 19–35.
- Murnane, Richard J., and B. Phillips. 1981. "What do Effective Teachers of Inner-City Children Have in Common?" *Social Science Research* 10, 1: 83–100.
- Murnane, R. J., and D. K. Cohen. 1986. "Merit Pay and the Evaluation Problem: Why Most Merit Pay Plans Fail and a Few Survive," *Harvard Educational Review* 56 (February): 1–17.
- Murnane, Richard J., and Randall Olsen. 1990. "The Effects of Salaries and Opportunity Costs on Length of Stay in Teaching," *Journal of Human Resources* 25: 106ff.
- Murnane, Richard J., Judith D. Singer, John B. Willett, James J. Kemple, and Randall J. Olsen. 1990. *Who Will Teach? Policies That Matter*. Cambridge: Harvard University Press.
- Oates, Wallace. 1969. "The Effects of Property Taxes and Local Public Spending on Property Values: An Empirical Study of Tax Capitalization and the Tiebout Hypothesis," *Journal of Political Economy* 77, 6 (November/December): 957–971.
- Raudenbush, Stephen W. 1988. "Educational Applications of Hierarchical Linear Models: A Review," *Journal of Educational Statistics* 13, 2 (Summer): 85–116.
- Ribich, Thomas I. 1968. *Education and Poverty*. Washington, DC: Brookings Institution.
- Ribich, T. I., and J. L. Murphy. 1975. "The Economic Returns to Increased Educational Spending," *Journal of Human Resources* 10, 1: 56–77.
- Roback, Jennifer. 1982. "Wages, Rents, and the Quality of Life," *Journal of Political Economy* 90, 6 (December): 1257–1278.
- Rosen, Harvey S., and D. J. Fullerton. 1977. "A Note on Local Tax Rates, Public Benefit Levels, and Property Values," *Journal of Political Economy* 85, 2 (April): 433–440.
- Sebold, F. D., and W. Dato. 1981. "School Funding and Student Achievement: An Empirical Analysis," *Public Finance Quarterly* 9, 1: 91–105.
- Tiebout, Charles M. 1956. "A Pure Theory of Local Expenditures," *Journal of Political Economy* 64 (October): 416–424.
- Toder, Eric J. 1972. "The Supply of Public School Teachers to an Urban Metropolitan Area: A Possible Source of Discrimination in Education," *Review of Economics and Statistics* 54, 4 (November): 439–443.
- Tron, Esther O. 1980. *Public School Finance Programs, 1978–1979*. Washington, DC: Government Printing Office.

- Underwood, Julie K. 1989. "Changing Equal Protection Analyses in Finance Equity Litigation," *Journal of Education Finance* 14 (Winter): 413-425.
- Walberg, Herbert J., and William J. Fowler, Jr. 1987. "Expenditure and Size Efficiencies of Public School Districts," *Educational Researcher* (October): 5-13.
- Walberg, Herbert J., and Sue Pinzur Rasher. 1974. "Public School Effectiveness and Equality: New Evidence and Its Implications," *Phi Delta Kappan* (September): 3-9.
- Wendling, Wayne. 1981. "Capitalization: Considerations in School Finance," *Educational Evaluation and Policy Analysis* (March-April).
- Wendling, Wayne, and Judith Cohen. 1986. "Education Resources and Student Achievement: Good News for Schools," *Journal of Education Finance* 7 (Summer): 44-63.
- Wachter, Kenneth W., and Miron L. Straf, Eds. 1990. *The Future of Meta-Analysis*. New York: Russell Sage.
- Wise, Arthur E. 1968. *Rich Schools, Poor Schools: The Promise of Equal Educational Opportunity*. Chicago: University of Chicago Press.

Can Schools Be Held Accountable for Good Performance?

A Critique of Common Educational Performance Indicators*

Robert H. Meyer
University of Chicago

Educational indicators are increasingly being used to assess the efficacy of American education. Local newspapers regularly report how students in local schools perform on nationally standardized tests, and a growing number of states publish formal school report cards that provide an assortment of student outcome, enrollment, and financial indicators. In April 1991, President Bush elevated the discussion of educational indicators to the national level with "America 2000," a proposal to establish a national examination system, complete with school district, state, and national report cards (U.S. Department of Education 1991).

The growing demand for educational performance indicators has been motivated in large part by a growing demand for public *accountability* defined in terms of hard outcomes, such as standardized test scores, rather than inputs, such as teacher qualifications, class size, and course requirements. Demands for public accountability have been particularly strong in states that have dramatically increased expenditures on education and in states that have launched major school improvement efforts. The increased demand for public accountability in elementary and secondary education has paralleled similar demands for increased accountability in other public sector activities, for example, the Job Training Partnership Act and the new JOBS program, enacted as part of the Family Support Act.

---

*This research has been supported by the La Follette Institute of Public Affairs and the Institute for Research on Poverty, University of Wisconsin-Madison. The author has benefited enormously from discussions with Bill Clune, Jeff Domnitz, and Andy Porter and from comments by Ken Bickers, Mike Wiseman, and participants in the LaFollette Institute public affairs workshop.

Despite the groundswell of interest in data on school performance, many educators and scholars fear that poorly implemented performance indicators could ultimately be worse than no indicators at all. These fears are not groundless. As will be discussed in this paper, performance indicators based on achievement tests could be flawed in two major ways. First, the achievement test underlying a performance indicator could be susceptible to “narrow” teaching to the test or could fail to reflect a school’s true educational objectives. Second, a performance indicator constructed from a simplistic or otherwise inappropriate statistical model could fail to reflect the true contribution of a school to growth in measured student achievement. Under these conditions, a high stakes system of educational performance indicators could severely distort the behavior of educators and students.

The purpose of this paper is to assess the statistical adequacy of the most commonly used educational performance indicators. One of the major conclusions of the analysis is that the typical indicators used to assess school performance—average and median test scores—are highly flawed as measures of school performance. As a result, they are of limited value, if not useless, for evaluating relative school performance or school performance over time. Indeed, simulation results indicate that changes over time in average test scores could very well be negatively correlated with actual changes in school performance.

The analysis also demonstrates that the typical indicators used to assess school performance are likely to provide schools with the perverse incentive to “cream,” that is, to raise measured school performance by educating only those students who tend to have high test scores. The potential for creaming is apt to be particularly strong in environments characterized by selective admissions. However, creaming could also exist in more subtle, but no less harmful, forms. For example, schools could create an environment that is relatively unsupportive for potential dropouts, academically disadvantaged students, and special education students, thereby encouraging these students to drop out or transfer to another school. Alternatively, high-quality teachers and administrators could gravitate to neighborhood schools that predominantly serve high-scoring students.

The paper is organized in nine major sections, the first of which is this introduction. The second section is a discussion of the problems that exist with traditional standardized tests; the third presents an

assessment of the validity of the average test score. I demonstrate that this commonly used indicator is highly flawed as an indicator of school performance. In the fourth section, I demonstrate that an alternative indicator, the gain indicator, avoids all but one of the major flaws associated with the average test score. In particular, the gain indicator fails to measure the value-added contributions of schools to growth in academic achievement. The seventh and eighth sections draw on simulated and actual data to illustrate the advantages of gain indicators over average summary scores. I first investigate value-added indicators, and then consider the consequences of evaluating schools on the basis of incomplete indicators. Finally, I present recommendations for the phased-in development of valid educational performance indicators. An appendix provides technical information concerning the simulations reported in the fifth section.

The Problems With Traditional Standardized Tests

Many educators and testing experts believe that there is a great need for new and improved ways of testing student achievement. A major problem with national standardized tests is that they are designed to appeal to all schools regardless of their educational objectives. These tests, if used in a high stakes indicator system, could drive teachers and administrators to focus almost exclusively on low-level academic content (Smith and O'Day 1990; Clune 1991). The achievement tests used as the basis for a performance indicator system should ideally reflect a balance of low- and high-end content so that the performance of schools that serve low- and/or high-achieving students can adequately be measured. This implies that a minimum competency test is unlikely to be satisfactory as the basis for measuring school performance. The problem with minimum competency exams is that many students receive a perfect score year after year. If the tests differ from one grade to the next, the recorded gain for these students is totally artificial. If the tests do not differ, their recorded gain is zero—in most cases, a vast understatement of their true gain in achievement. The simple achievement models presented later in this paper are not really appropriate for tests that exhibit low ceilings and/or high floors. However, the models

could be extended to allow for the “censoring” of test scores at the high and low extremes of the test score distribution.

Critics of standardized tests also argue that conventional multiple choice tests are not well suited to assessing skills involving higher order thinking and problem solving, the kinds of skills that are increasingly valued in our economy. They argue that the multiple choice format is generally limited to asking simple questions that have definite answers. As a result, a history exam is reduced to questions about dates and events, rather than the causes of the Civil War; a mathematics exam is reduced to a long series of addition and multiplication problems, rather than questions involving the application of mathematics to solving real-world problems. It is feared that a system of performance indicators based on such tests is likely to encourage teachers and administrators to focus their teaching on repetitive, rote learning.

These criticisms have stimulated a number of states to begin developing new, performance-based tests (Dominitz and Meyer 1991). One commonplace example of an authentic performance-based test is the field portion of a driving test. A driving test assesses, more or less, what a driver needs to know to drive on city streets. Indeed, the best way to pass a driving test is to practice driving. In contrast, typical standardized math tests fail to assess what most students need to know about mathematics, the capacity to tackle extended real-world problems calling for the application of diverse mathematics skills. Advocates of performance-based tests argue that these tests will be relatively immune to the phenomenon of narrow teaching to the test and more congruent with state educational curriculum goals.

Level Indicators

Standardized student testing is conducted for a variety of different reasons: to provide information on individual students and obtain aggregate school-level indicators. At the student level, for example, standardized test scores may be used to diagnose student strengths and weaknesses in subskill areas,¹ to guide teachers in providing instruction that matches the needs of individual students, to guide students in making curriculum and career choices, to determine, in states that have

minimum competency examinations, whether students are eligible for graduation, and to guide postsecondary institutions and employers in making admissions and hiring decisions, respectively.

These data, if aggregated to the classroom or school level, yield educational indicators that measure, for example, the share of students scoring above or below certain thresholds or the average level of achievement. I refer generally to statistics of this kind as level indicators. As previously mentioned, level indicators are widely reported by schools. Indeed, they are calculated and readily made available by the companies that provide testing services to schools throughout the nation (Goldman 1990). They are also reported at the national level by the National Assessment of Educational Progress. Unfortunately, some of the level indicators reported by schools and states are subject to obvious statistical flaws. Well-known examples include average SAT and ACT scores. The problem with these indicators is that they are based on nonrandomly selected groups of students—in particular, those students who aspire to attend selective colleges or universities. As discussed by Hanushek and Taylor (1990), Powell and Steelman (1984), and Wainer (1986), these indicators tend to be highly unreliable as measures of the true level of achievement in schools and states. In this paper, I limit my analysis to level indicators that are not subject to these problems.

If correctly constructed and based on appropriate tests, level indicators convey potentially useful *descriptive* information concerning the proficiency levels of students in particular classrooms or schools. Indeed, they could sensibly be used to target assistance (financial or otherwise) to schools that serve students with low test scores. The critical question for the present discussion is whether such indicators are valid and useful measures of school or classroom performance. The answer to this question is no. School performance indicators, by definition, must validly measure the contribution of schools to growth in student achievement for students in particular grades or sequence of grades.

Average (or median test) scores fail to do this for four reasons. First, the average test score fails to *localize* school performance to a specific classroom or grade level—the natural unit of accountability in a traditional school.² This lack of localization is, of course, most severe at the highest grade levels. In my judgment, a performance indicator that

fails to localize school performance to a specific grade level or classroom is likely to be a relatively weak instrument of public accountability.

Second, the average test score reflects information that is *aggregated across time and grade levels* and therefore tends to be grossly *out of date*. For example, consider the average test score for a group of high school seniors. The test scores for these students reflect learning that occurred in kindergarten, roughly twelve-and-one-half years earlier, through the twelfth grade. Indeed, a twelfth-grade level indicator could be dominated by information that is ten or more years old.³ The fact that average test scores reflect out-of-date information severely weakens them as instruments of public accountability. In order to allow educators to react in a timely and responsible fashion, performance indicators must reflect information that is current.

Third, average test scores at the school, district, and state levels tend to be highly *contaminated* due to student mobility in and out of different school systems. For example, the typical twelfth-grade student is likely to attend several different schools over the period spanning kindergarten through twelfth grade. For this student, a test score reflects the contributions of more than one and possibly many different schools. The problem of contamination is compounded by the fact that rates of student mobility tend to differ dramatically across schools. Contamination is apt to be especially high in communities that undergo rapid population growth or decline or experience significant changes in their occupational and industrial structure. Contamination due to student mobility is probably a relatively minor problem at the national level, since rates of in- and out-migration are low compared to rates of mobility within the nation.

Fourth, the average test score is not a *value-added* indicator; that is, it fails to measure the distinct contribution of a school to growth in educational achievement. As a result it absorbs differences across schools in student achievement levels that are due not to differences in school productivity but rather to variations in student achievement prior to entering school and to differences in growth in student achievement that are systematically related to differences in student and family background characteristics.

In summary, the average test score suffers from four major flaws, any one of which could be sufficient to invalidate it as a measure of

school performance. In the next section I therefore consider an alternative indicator that largely avoids the problems of nonlocalization, aggregation across time and grade levels, and contamination, namely, the gain indicator. Immediately following is a series of simulations that compare the average test score relative to the gain indicator.

Gain Indicators

The gain indicator measures the average growth (or gain) in achievement from one point in time to another for a given cohort of students. If students are tested at least once a year, the gain indicator largely avoids three of the problems that seriously undermine the average test score as a valid and up-to-date measure of school performance: the problems of nonlocalization, aggregation across time and grade levels, and contamination due to student mobility. However, the gain indicator does not measure the value-added contribution of schools to growth in student achievement, that is, it does not measure school performance. Rather, it measures the joint contributions of students, families, communities, and schools to growth in student achievement. As such, it is an extremely informative *descriptive* indicator that should be included, along with the value-added indicators introduced below, in a comprehensive system of educational indicators.

The quality of the gain indicator depends critically on the frequency of student testing. Annual (or more frequent) testing is ideal for several reasons. First, performance is localized to single grade levels, the natural unit of accountability. Second, the information reflected in the indicator is completely up to date. Third, contamination due to student mobility is limited only to students who transfer schools during the school year.

As the time interval between tests increases, the problems of localization, contamination, and aggregation over time and grade levels become more acute. In fact, for time intervals of more than two years, it could prove difficult to construct valid and reliable gain indicators for schools with high mobility rates. There are two options in such cases. First, mobile students could simply be excluded from the data for a classroom or school. Dyer, Linn, and Patton (1969) refer to this as

the “matched sample” approach. The problem with this approach is that nonmobile students are apt to be unrepresentative of the school population as a whole, both in terms of student characteristics and educational experiences. Moreover, the number of nonmobile students in such cases could be simply too small to yield reliable (statistically precise) estimates of average student gain. The second option is to include mobile students in the gain comparison for a given school even though the students obtained part of their schooling from another school. Of course, this option is feasible only if mobile students take the same tests in different schools and if their test scores are made available to the schools to which they move or exit. This clearly would be feasible only in states that have mandated state assessment systems. Even so, students who move across state lines would be lost unless the states happen to use the same state tests and are prepared to exchange student test data. A more fundamental problem with this approach is that the contamination introduced by mobile students severely jeopardizes the validity of the gain indicator if the mobility rate is high. The bottom line is that infrequent testing seriously compromises the validity of the gain indicator.

How Bad is the Average Test Score as a Measure of School Performance? Simulation Results

This section presents a series of simulations designed to assess whether the average test score has *any* value as a measure of educational productivity. I consider the validity of the average test score for comparisons across schools and for comparisons over time for the same school. The second type of comparison is particularly relevant for the purposes of evaluating the efficacy of school reform efforts.

Let $L(c, g)$ represent the average level of achievement in a particular school for cohort c at the end of grade g . Similarly, let $G(c, g)$ represent the average gain in achievement in a particular school for cohort c from the end of grade $(g-1)$ to grade g , that is,

$$G(c, g) = L(c, g) - L(c, g-1). \quad (1)$$

Equation (1) implies that, for a given cohort, the average level of achievement at the end of a particular grade, say grade 10, is the sum of prior gains in achievement plus the initial average level of achievement, that is,

$$L(c, 10) = L(c, 0) + G(c, 1) + \dots + G(c, 10). \quad (2)$$

Given alternative assumptions concerning initial achievement and the pattern of gain values over time and across grade levels, I can compute the average level of achievement at the end of grade 10 for each cohort.

To emphasize the contrast between average achievement and the gain in achievement, I assume that average initial achievement and average student characteristics are identical for all schools at all points in time.⁴ I also assume, for simplicity, that all students begin first grade at the same age and advance from one grade to the next each year. In this case, a unique time index is implied by the cohort and grade level indices. The relationship between time, birth cohort, and grade level is given by the formula⁵

$$t = c + g + 6.$$

To facilitate comparisons across schools at the same grade level, I standardize the school gain values so that the average gain for the entire population at a given point in time is equal to zero at each grade level. Average 10th grade achievement is similarly standardized to have mean zero. Finally, I assume that the achievement test underlying this analysis is scaled so that the standard deviation of school gain values is approximately equal to 10 in the typical grade.⁶ To provide the reader with some intuitive sense of the standardized gain values, table 1 lists percentile values associated with a range of gain indicator values.

The first pair of simulations illustrate the failure of average test scores to localize school performance to specific grade levels. Subsequent simulations illustrate the consequences of contamination and aggregation across time and grades. Technical details of the simulations are presented in the appendix.

Table 1. Gain Percentile Values, Given the Assumption that Average Gains are Normally Distributed

Gain indicator values, given zero mean and standard deviation equal to 10	Gain percentile values, given the assumption that average gains are normally distributed
30	99.9
20	97.9
10	84.1
5	69.2
0	50.0
-5	30.8
-10	15.9
-20	2.3
-30	0.1

The first simulation, as summarized in table 2, contrasts three schools that differ in terms of their patterns of (standardized) gain in grades one through six and grades seven through ten, respectively. To simplify the analysis I assume that these patterns persist over time and that there is no student mobility. School 1 exhibits gain values of zero (the average) at all grade levels. School 2 exhibits exceptionally high gain values in the upper grades and exceptionally low gain values in the lower grades. Finally, school 3 exhibits a pattern of gain values that is exactly opposite to the pattern exhibited for school 2. As indicated, the three schools differ fundamentally in terms of their gain values in the early and late grades. Despite these differences, however, the schools are indistinguishable in terms of their average level of achievement at the end of tenth grade. The exceptionally high and the exceptionally low gain values simply cancel out for schools 2 and 3.

A similar result is observed in the second simulation, as depicted in figure 1. Figure 1 charts the average level of tenth-grade achievement over time, prior to and after the implementation of hypothetical academic reforms in 1992. The academic reforms are assumed to follow an era of stable but average gains in achievement at all grade levels. Panels A and B in figure 1 depict two different scenarios. In panel A the average achievement gains at each grade level increase gradually after 1991. In panel B, the average achievement gains also increase

steadily, but they are limited to grades seven to ten. The gain values are indicated on the graph by the solid gray lines. The tenth-grade achievement levels are indicated on the graph by the solid black lines. As in the previous simulation, the two schools differ substantially in terms of their gain values at different grade levels. Despite these differences, however, there is no perceptible difference between the two schools in terms of average tenth-grade achievement. In short, these two simulations demonstrate that average test scores provide no information on differences in productivity between different levels of a school system. They do, however, suggest that average test scores provide at least a rough indication of the productivity of the school system overall. In fact, this is generally not true, as is demonstrated below.

Table 2. Average Tenth-Grade Achievement by School, Given Alternative Patterns of Gain

School	Average gain per grade			Average achievement at the end of tenth grade
	Initial achievement	Grades 1 to 5	Grades 6 to 10	
1	0	0	0	0
2	0	20	-20	0
3	0	-20	20	0

The second set of simulations illustrates the problem of aggregation across time and grade levels. These simulations demonstrate vividly how average test scores are determined in large part by past gains in achievement and hence are apt to be quite misleading as indicators of current gains. To highlight the problem of aggregation across time and grade levels I assume that achievement gains within a school are identical at all grade levels and that there is no student mobility. Figure 2 charts average tenth-grade achievement and average achievement gains over time, prior to and after the introduction of hypothetical academic reforms in 1992. Panel A of figure 2 depicts a scenario in which academic reforms reverse a trend of gradual deterioration in average achievement gains across all grades and initiate a trend of gradual improvement in average achievement gains across all grades. Panel B

Figure 1. Average Tenth-Grade Achievement Given Alternative Patterns of Gain in Grades 1 to 6 and 7 to 10

Figure 2. Average Tenth-Grade Achievement Given Alternative Patterns of Gain Over Time

of Figure 2 depicts a scenario in which academic reforms have absolutely no effect on average achievement gains. The reforms, however, are preceded by an era of gradual deterioration in average achievement gains across all grades, followed by a brief period (1987–1991) of gradual improvement across all grades. As indicated in the graph, the average tenth-grade test score provides a totally misleading view of the effectiveness of the hypothetical academic reforms implemented in 1992. In panel A, the average 10th grade test score *declines* for five years after the introduction of successful reforms. In panel B, the average tenth-grade test score *increases* for a decade after the introduction of reforms that have no effect on student achievement growth. These results are admittedly somewhat counter intuitive. They arise from the fact that 10th grade achievement is the product of gains in achievement accumulated over a ten-year period.⁷ The noise introduced by this type of aggregation is inevitable if school performance is at all variable over time. (The interested reader may want to peruse appendix tables A-3 and A-4. These tables provide additional information concerning the two simulations discussed above.)⁸

The problem of aggregation over time and grade levels also introduces noise into the comparisons of different schools at the same point in time. The degree to which noise of this type affects the relative ranking of schools depends on whether the variance over time in average achievement growth is large relative to the variance across schools in achievement growth. To illustrate this point, figure 3 considers the consequences of aggregation over time and grade levels for two schools that are identical in terms of average achievement gains over the long term. In the short term, however, average achievement gains are assumed to vary cyclically. For school 1, average gains alternate between ten years of gradual decline and ten years of gradual recovery. For school 2, average gains alternate between ten years of gradual improvement and ten years of gradual decline. These patterns are depicted in panel B of figure 3. The *correct* ranking of schools, based on average achievement growth, is noted in the graph. Panel A depicts the associated levels of average tenth-grade achievement for the two schools. The ranking of schools based on this indicator is also noted. The striking aspect of figure 3 is that the average tenth-grade test score ranks the two schools incorrectly exactly 50 percent of the time. In short, the noise introduced by aggregation over time and grade levels is

Figure 3. Average Tenth-Grade Achievement Given Alternative Cycles of Decline and Recovery in Average Gain

Note: Highest ranked school, according to this indicator, is indicated by the number in the black square.

Note: The correct ranking of schools, as determined by average gain, is indicated by the number in the black square.

particularly troublesome if one is comparing schools that are roughly comparable in terms of long-term average achievement growth. On the other hand, this problem is less serious for schools that differ dramatically in terms of long-term average achievement growth. It is also less serious if cycles of decline and improvement tend to be perfectly correlated. This seems unlikely as a general rule.

The third and final set of simulations illustrates the possible consequences of contamination due to student mobility. These simulations illustrate the extreme sensitivity of average test scores to in-migration of students. To highlight the consequences of student mobility I assume that achievement gains within a school are identical at all grade levels and over time. The first simulation envisions an environment in which there are three types of schools that vary in terms of their average achievement growth.⁹ Panel A of table 3 reports the effects on average 10th grade achievement of alternative rates of student mobility among the three schools. Panel B of table 3 reports the fraction of students who change schools, given alternative annual rates of student mobility. Notice that student mobility causes average tenth-grade test scores to collapse toward zero, the average level. For the high- and low-gain schools, for example, an annual mobility rate of 20 percent leads to a reduction in average test scores of over 70 percent. In other words, the average test score is severely biased against high gain schools that happen to serve highly mobile student populations. These numbers suggest that average test scores are apt to be highly misleading indicators of school quality for schools exposed to high rates of student mobility.¹⁰

If rates and patterns of student mobility vary over time, average test scores are also apt to provide a misleading picture of actual changes in school quality over time. This point is illustrated in figure 4, which simulates the effects on average tenth-grade achievement of an influx of students from a low-quality to a high-quality school system. Events of this kind undoubtedly occur frequently throughout the nation as school systems merge, communities grow, and the occupational structure of jobs evolve in a local labor market. Panel A of figure 4 simulates the effects of a gradual influx of students that takes place over a ten-year period: 1992–2001. Panel B simulates the effects of an instant influx of students in 1992. Despite the fact that average achievement growth remains constant after the influx of students, average achieve-

ment levels decline precipitously following the influx of students under either scenario. In the case of the gradual influx of students, the average level of achievement declines by as much as 50 percent. Moreover, average achievement does not return to its 1991 level until the year 2010. In the case of the instant influx of students, the average level of achievement falls instantly by 90 percent and is back to its 1991 level within a decade. In short, idiosyncratic shifts in patterns of student mobility have the potential to grossly contaminate the average test score as an indicator of contemporaneous school performance.

Table 3. Consequences of Student Mobility

A. Average Tenth-Grade Achievement by School, Given Alternative Rates of Student Mobility						
Gain value	Annual mobility rate (percent)					
	0	2	5	10	20	40
High	100	86.9	74.9	56.7	26.8	13.4
Medium	0	4.4	3.4	0.3	-3.8	-2.3
Low	-100	-86.7	-69.3	-56.7	-28.1	-11.3
B. The Fraction of Students Who Change Schools while in Grades 1 through 10, Given Alternative Rates of Student Mobility (percent)						
One or more changes	0	17.0	37.0	62.7	89.3	99.7
Two or more changes	0	1.7	8.7	21.7	56.0	94.0

The simulations presented in this section demonstrate that average test scores have the potential to provide a totally misleading portrait of educational productivity, both over time and across schools. Indeed, the simulations possibly understate the degree to which average test scores are flawed as valid measures of school performance since they address the problems of nonlocalization, aggregation, and contamination one at a time, not simultaneously. Fortunately, gain indicators largely avoid the three problems investigated in the above simulations. Moreover, these indicators are generally easy to compute.

Figure 4. Average Tenth-Grade Achievement Given Different Patterns of Student Mobility

An Example Based On National Data

The policy significance of the above discussion is aptly illustrated using data on average mathematics scores from 1973 to 1986 from the National Assessment of Education Progress (NAEP). As indicated in panel A of table 4, NAEP scores for eleventh graders exhibit the by-now familiar pattern of sharp declines from 1973 to 1982 and then partial recovery between 1982 to 1986. The eleventh-grade data, by themselves, are fully consistent with the premise that academic reforms in the early and mid-1980s generated substantial gains in academic achievement. In fact, an analysis of the data based on gain indicators rather than average test scores suggests the opposite conclusion—see panel B of table 4. Gain indicators were constructed in panel B by computing the change in average test scores over time for given birth cohorts.¹¹ The gain indicators reveal that achievement growth during the 1982 and 1986 period was actually no better than achievement growth during the prior 1978 to 1982 period. In fact, gains from seventh to eleventh grade were actually slightly lower during the 1982 to 1986 period than in previous periods! The rise in eleventh-grade math scores from 1982 to 1986 apparently stems from an earlier increase in achievement growth for that cohort rather than from an increase in achievement growth over grades seven to eleven. In short, these data provide no support for the notion that high school academic reforms generated significant increases in test scores during the mid-1980s. These data also vividly confirm the general superiority of gain indicators, relative to level indicators, as measures of educational productivity.

Value-Added Indicators

As discussed in the previous section, the gain indicator measures the joint contribution of students, families, communities, and schools to growth in student achievement. The problem is that a school may rate highly in terms of a gain indicator primarily or solely because the school serves students capable of rapid achievement growth. Unfortu-

nately, failure to achieve a valid measurement of school performance could provide schools with the incentive to improve “measured” performance simply by trying to control the types of students who attend their schools.

Table 4. NAEP Mathematics Exam Data

Grade/Age	A. Average Test Scores			
	1973	1978	1982	1986
3rd/9	219.1	218.6	219.0	221.7
7th/13	266.0	264.1	268.6	269.0
11th/17	304.4	300.4	298.5	302.0
	B. Average Test Score Gains			
	1973 to 1978	1978 to 1982	1982 to 1986	
3rd to 7th/9 to 13	45.0	50.0	50.0	
7th to 11th/13 to 17	34.4	34.4	33.4	

SOURCE: Dossey et al (1988).

In order to isolate the distinct contribution of a school to growth in student achievement, a statistical model must be used. The statistical model, if valid, allows one to estimate for each school or classroom the expected (or average) gain in achievement that would be realized by a given student. In this sense, the model estimates school performance *controlling* for differences across schools in student characteristics and perhaps school-level variables such as aggregate student and community characteristics. If these characteristics differ significantly across schools or classrooms, value-added and gain indicators could differ significantly. Dyer, Linn, and Patton (1969), Hanushek (1972), and Murnane (1975) were among the first researchers to estimate value-added indicators of school performance.

What variables should be included as control variables in a value-added model of student achievement and school performance? From the perspective of school accountability, it is important to control for all factors external to schools, in particular, student and community characteristics. Performance with respect to intrinsic school and classroom factors is what matters. In practice, most school districts have ready access to some, but not all, of the student characteristics that are likely to determine student achievement: (1) Is a student eligible for a

free or reduced-price school lunch? (2) Is a student eligible for special education services? (3) Does a student's family receive financial assistance from welfare programs? and (4) Is the student classified as being at-risk? It is not well known whether these variables adequately control for differences across schools in average student characteristics. If not, value-added indicators, as implemented, might not fully eliminate the distortions (see below) associated with level and, to a lesser extent, gain indicators.

The exact relationship between a gain and value-added indicator is as follows. For a given cohort at a given grade level, the average gain in student achievement G is the sum of two terms: the value-added contribution of a school to growth in student achievement P and the average contribution of (external) student and community characteristics to growth in student achievement $F(X)$, where X represents a set of student and community characteristics, and the function F is estimated from an appropriate statistical model of student achievement growth.¹² Similarly, a level indicator is the sum of three terms: P , $F(X)$, and average achievement prior to entering a given grade (see above section on simulation results). From the perspective of measuring school performance, the term $F(X)$ is a source of error in a gain and level indicator. Prior average achievement is an additional source of error in a level indicator.

The fact that gain and level indicators measure school performance with error has important implications for the use of these indicators for purposes of school choice and accountability. Because of the contamination due to these errors, level indicators, and to a lesser extent gain indicators, are likely to give students the wrong signals about which schools to attend. In practice, this means that prospective students, both academically advantaged and disadvantaged, could be fooled into abandoning an excellent neighborhood school simply because the school served students that were disproportionately academically disadvantaged. At the other extreme, these indicators could contribute to complacency on the part of families whose children attend schools that disproportionately serve academically advantaged students. In fact, these schools could be adding relatively little to the achievement growth of their students. In short, indicators other than the value-added performance indicator convey potentially inaccurate information about school quality and therefore are likely to distort the school choices of

students and families. As a result, student achievement is apt to be lower than it would otherwise be.

The consequences of using invalid performance indicators for purposes of public accountability are if anything potentially much worse than in the case of school choice. This stems from the fact that the indicators used for purposes of public accountability have the potential to influence, if not determine, the objectives of teachers and administrators. Indeed, if teachers and administrators are in any way rewarded or penalized on the basis of their performance with respect to a given indicator, they are likely to respond to these incentives by trying to improve their measured performance. In other words, they will have an incentive to "teach to the test." More to the point, they will have an incentive to "teach to the indicator derived from the test."

This phenomenon is the key to understanding why valid performance indicators are potentially capable of generating substantial genuine improvements in school quality. However, it is also the key to understanding how statistically invalid indicators could severely distort the behavior of teachers and administrators. Consider, for example, the consequences of using a level indicator to evaluate school performance. A level indicator is the sum of school performance and two error components that are determined by average student characteristics, average prior achievement, and community characteristics. If this indicator is used to evaluate school performance, it provides teachers and administrators with the incentive to raise measured school performance by teaching only those students who rate highly in terms of average student characteristics, average prior achievement, and community characteristics. In general, these students will be high socioeconomic status, academically advantaged students. This is the phenomenon referred to earlier as "creaming".

The potential for creaming is apt to be particularly strong in environments where schools have the authority to admit or reject prospective students and to expel already enrolled students. However, the problem could also exist in more subtle but no less harmful forms. For example, schools could: (1) create an environment that is relatively inhospitable to academically disadvantaged students, (2) provide course offerings that predominantly address the needs of academically advantaged students, (3) fail to work aggressively to prevent students from dropping out of high school, (4) err on the side of referring "prob-

lem" students to alternative schools, (5) err on the side of classifying students as special education students (if these students are exempted from statewide testing), and (6) make it difficult for low-scoring students to participate in statewide examinations. These activities are all designed to improve average test scores in a school, not by improving school quality but by selecting high-scoring students.

As an alternative to trying to select high-scoring students, high-quality teachers and administrators could gravitate to neighborhood schools that predominantly serve high-scoring students. Hence, using the average test score as a high-stakes performance indicator could trigger an exodus of highly skilled educators from schools that disproportionately serve academically disadvantaged students.

One final problem with the average test score is that teachers, administrators, and the public are apt to correctly perceive it as an unfair measure of school performance, thereby undermining the legitimacy of the entire indicator system. Indeed, there is some evidence that this has occurred in one of the states studied by Dominitz and Meyer (1991).

The criticisms discussed above apply equally, although with less force, to the gain indicator, since it is subject to a single source of error, $F(X)$.

Multiple Dimensions Of Performance

Thus far, I have ignored the fact that schools typically have multiple objectives, both academic and nonacademic. Several issues that arise in the context of multiple objectives need to be addressed at this point. First, it seems likely that an ideal performance indicator system would include separate indicators designed to match each and all of the objectives adopted by a school. Such a system would probably include indicators designed to measure school performance in conventional academic subjects, possibly mathematics, science, literature, history, reading, and writing; but it could also include indicators of school performance in other areas, for example, citizenship, employment readiness, and fine arts. The problem is that it could prove technically difficult, burdensome, and expensive to measure outcomes in all of

these areas. If indicators are available for only a subset of objectives, however, it is possible, even likely, that those objectives would effectively dominate all other objectives. This could distort the behavior of teachers and administrators by giving them the incentives to devote most of their instructional time to the subjects covered by performance indicators.

One solution to this dilemma is to measure school performance in the areas that are considered to be central to the missions of schools. Indeed, there could be advantages to adopting a more limited set of educational objectives than currently exists. The adoption of performance indicators could conceivably force parents and educators to decide what educational objectives are really important.

It seems inevitable, though, that some important educational objectives could be too difficult to measure. If so, one alternative is to measure the inputs (instructional time and resources) devoted to these activities. This could counteract the incentives to limit instruction in these activities in order to devote more time to activities that are evaluated. On the other hand, the absence of performance indicators in particular areas eliminates the opportunity to hold schools accountable for their performance in these areas.

Second, it seems likely that some educational objectives could be more important than others. How can priorities of this nature be incorporated into an indicator system? One possibility is to construct an overall performance indicator that reflects the preferences of an individual, community, or state. A linear, weighted average of individual performance indicators is one particularly simple example of a preference function. Such a system has recently been adopted in California (Dominitz and Meyer 1991). One potential weakness of the linear preference function is that it allows high performance in one dimension to substitute fully for low performance in another. In fact, most students and parents are likely to prefer schools that are very good in many dimensions, as opposed to schools that are excellent in some areas, poor in others. If so, states and communities could adopt preference functions that limit the degree of substitutability between competing objectives. Examples of such functions include the Cobb-Douglas and constant elasticity of substitution (CES) functions (Henderson and Quandt 1971). This is clearly an area for further research.

Conclusions and Recommendations

The average test score, one of the most commonly used indicators in American education, is highly suspect as an indicator of school performance.¹³ This indicator suffers from four major deficiencies: it fails to localize school performance to the classroom or grade level; it aggregates information on school performance across time and grade levels; it is contaminated by student mobility; and it fails to measure the distinct contribution of schools to growth in student achievement. As a result, the average test score is a weak, if not counterproductive, instrument of public accountability. The gain indicator, on the other hand, avoids three of the four problems that plague the average test score. As such, it is a very useful descriptive indicator. The value-added indicator has the major advantage that it avoids all four of the problems that affect level indicators. In particular, it eliminates the incentive for schools to cream.

The value-added approach to measuring school performance relies on a statistical model to identify the distinct contributions made by schools to growth in student achievement. The quality of a value-added indicator is determined by four factors: the frequency with which students are tested, the quality and appropriateness of the tests that underlie the indicators, the adequacy of the control variables included in the appropriate statistical models, and the technical quality of the statistical models used to construct the indicators.

In terms of the first issue, I believe that states need to seriously consider testing students at every grade level, as is currently done in South Carolina (Dominitz and Meyer 1991), or at least at every other grade level, beginning with kindergarten. Annual testing maximizes accountability by localizing school performance to the most natural unit of accountability, the grade level or classroom. It also limits the contamination caused by student mobility and yields up-to-date information on school performance. Less frequent testing, for example, testing at grades kindergarten, four, eight, and twelve, might be acceptable for national purposes, since student mobility is not really at issue at the national level.¹⁴ For purposes of evaluating local school performance, however, the problems created by student mobility argue strongly for frequent testing. To limit the costs and burden imposed by frequent stu-

dent tests, however, it might be sensible to vary the frequency of testing across schools. Annual testing could be implemented only in schools or school districts where student mobility is high. In addition, annual testing could be implemented in areas with limited enrollments in order to improve the reliability of estimates in these areas, and in schools with low measured performance in order to monitor these schools with greater vigilance.

With respect to the second and third issues, it is important that states make it a major priority to collect extensive and reliable information on student and family characteristics and to develop state tests that are technically sound and fully attuned to their educational goals. Finally, further research is needed to assess the sensitivity of estimates of school performance indicators to alternative statistical models.

NOTES

1. For diagnostic purposes student test scores are often reported separately by subskill areas.
2. This point also applies to classrooms that serve students in more than one grade and ungraded classrooms.
3. This would occur, for example, if the variability over time of school performance is higher in elementary school than in middle or high school.
4. This assumption guarantees that differences across schools in average gain reflect differences in school performance rather than differences in student characteristics.
5. For example, the cohort born in 1980 entered first grade in 1986 and is expected to complete twelfth grade in 1998. Note that the concept of the birth cohort needs to be modified slightly to accommodate school districts that require first graders to be six years old *prior* to beginning school.
6. To further facilitate comparisons across schools at the same grade level, gain values could be standardized so that the standard deviation is equal to ten for every grade in every year. The disadvantage of this approach is that gain indicators constructed in this fashion are not comparable across grades or over time.
7. In the simulations discussed in the text, the average tenth-grade test score is, in fact, exactly equal to a ten-year moving average of average achievement gains. This stems from the simple assumption that achievement gains are identical at different grade levels in the same year.
8. The appendix tables report achievement gains by grade level and cohort. As indicated in the text, achievement gains change from year to year but are always identical across different grade levels in the same year. This shows up in appendix tables A-3 and A-4 as gain values that are equal on diagonal lines that run from the bottom left to the top right of the tables.
9. Average growth is assumed to be equal to 10, 0, and -10, respectively, in the three types of schools. See appendix A for additional details.
10. This conclusion is based on the assumption that at least some student mobility occurs across schools of different quality, a reasonable supposition, we think, in the absence of contrary data.

11. NAEP was originally designed to permit this type of analysis. In mathematics, the tests have generally been given every four years at grade levels spaced four years apart. For this illustrative analysis, I assume that average test scores in 1973 are comparable to the unknown 1974 scores.

12. For concreteness, consider the following statistical mode of achievement growth for students in a given grade:

$$Y(i,j) = a(j) + \sum_k b(k) X(i,j,k) + e(i,j)$$

where i and j index individuals and schools respectively, Y represents growth in student achievement, X represents a set of student and community characteristics (indexed by k), $a(j)$ represents a school-specific intercept, b represents a set of coefficients (indexed by k), and e represents a random error term. The gain indicator for school j is given by $G = \sum Y(i,j)/n(j)$, where $n(j)$ = the number of students in school j . The value-added performance indicator for school j is given by $P = a(j)$. The average contribution of external characteristics in school j is given by

$$F(X) = \sum_i \sum_k b(k) X(i,j,k)/n(j).$$

13. Other level indicators, such as the median test score, are similarly suspect.

14. A kindergarten test is needed so that the growth in student achievement in grades one through four can be monitored. In our view, the National Assessment of Educational Progress and recent proposals for national testing in grades four, eight, and twelve are seriously flawed by their failure to include a test at the kindergarten or first-grade level. I suspect that one reason for this omission is that both enterprises are insufficiently aware of the flaws of level indicators and insufficiently aware of the advantages of gain and value-added indicators.

Appendix

Descriptions Of Reported Simulations

This appendix presents results for the simulations presented in the text. Each simulation is defined in terms of the gain in achievement accrued by a student at a particular school in a given grade at a given point in time. The birth cohort subscript is implied by the grade and time subscripts, as discussed in the text. It is given by $c = t - g - 6$. For simplicity, I assume that students begin first grade at age six and advance to subsequent grades one year at a time. Gains in achievement are reported by grade and cohort and tenth-grade achievement for some of the simulations. Gains in achievement for a given year are reported on diagonal lines that run from the bottom left to the top right of the tables.

Appendix Table A-1. Data for Figure 1A

Year cohort completes grade 10	Average gain by grade										Average achievement in grade 10
	1	2	3	4	5	6	7	8	9	10	
1987	0	0	0	0	0	0	0	0	0	0	0
1988	0	0	0	0	0	0	0	0	0	0	0
1989	0	0	0	0	0	0	0	0	0	0	0
1990	0	0	0	0	0	0	0	0	0	0	0
1991	0	0	0	0	0	0	0	0	0	0	0
1992	0	0	0	0	0	0	0	0	0	1	1
1993	0	0	0	0	0	0	0	0	1	2	3
1994	0	0	0	0	0	0	0	1	2	3	6
1995	0	0	0	0	0	0	1	2	3	4	10
1996	0	0	0	0	0	1	2	3	4	5	15
1997	0	0	0	0	1	2	3	4	5	6	21
1998	0	0	0	1	2	3	4	5	6	7	28
1999	0	0	1	2	3	4	5	6	7	8	36
2000	0	1	2	3	4	5	6	7	8	9	45
2001	1	2	3	4	5	6	7	8	9	10	55

Appendix Table A-2. Data for Figure 1B

Year cohort completes grade 10	Average gain by grade										Average achievement in grade 10
	1	2	3	4	5	6	7	8	9	10	
1987	0	0	0	0	0	0	0	0	0	0	0
1988	0	0	0	0	0	0	0	0	0	0	0
1989	0	0	0	0	0	0	0	0	0	0	0
1990	0	0	0	0	0	0	0	0	0	0	0
1991	0	0	0	0	0	0	0	0	0	0	0
1992	0	0	0	0	0	0	0	0	0	2	2
1993	0	0	0	0	0	0	0	0	2	4	6
1994	0	0	0	0	0	0	0	2	4	6	12
1995	0	0	0	0	0	0	2	4	6	8	20
1996	0	0	0	0	0	0	4	6	8	10	28
1997	0	0	0	0	0	0	6	8	10	12	36
1998	0	0	0	0	0	0	8	10	12	14	44
1999	0	0	0	0	0	0	10	12	14	16	52
2000	0	0	0	0	0	0	12	14	16	18	60
2001	0	0	0	0	0	0	14	16	18	20	68

Appendix Table A-3. Data for Figure 2A

Year cohort completes grade 10	Average gain by grade										Average achievement in grade 10
	1	2	3	4	5	6	7	8	9	10	
1981	19	18	17	16	15	14	13	12	11	10	145
1982	18	17	16	15	14	13	12	11	10	9	135
1983	17	16	15	14	13	12	11	10	9	8	125
1984	16	15	14	13	12	11	10	9	8	7	115
1985	15	14	13	12	11	10	9	8	7	6	105
1986	14	13	12	11	10	9	8	7	6	5	95
1987	13	12	11	10	9	8	7	6	5	4	85
1988	12	11	10	9	8	7	6	5	4	3	75
1989	11	10	9	8	7	6	5	4	3	2	65
1990	10	9	8	7	6	5	4	3	2	1	55
1991	9	8	7	6	5	4	3	2	1	0	45
1992	8	7	6	5	4	3	2	1	0	1	37
1993	7	6	5	4	3	2	1	0	1	2	31
1994	6	5	4	3	2	1	0	1	2	3	27
1995	5	4	3	2	1	0	1	2	3	4	25
1996	4	3	2	1	0	1	2	3	4	5	25
1997	3	2	1	0	1	2	3	4	5	6	27
1998	2	1	0	1	2	3	4	5	6	7	31
1999	1	0	1	2	3	4	5	6	7	8	37
2000	0	1	2	3	4	5	6	7	8	9	45
2001	1	2	3	4	5	6	7	8	9	10	55

Appendix Table A-5. Data for Figure 3, School 1

Year cohort completes grade 10	Average gain by grade										Average achievement in grade 10
	1	2	3	4	5	6	7	8	9	10	
1981	2	4	6	8	10	12	14	16	18	20	110
1982	4	6	8	10	12	14	16	18	20	18	126
1983	6	8	10	12	14	16	18	20	18	16	138
1984	8	10	12	14	16	18	20	18	16	14	146
1985	10	12	14	16	18	20	18	16	14	12	150
1986	12	14	16	18	20	18	16	14	12	10	150
1987	14	16	18	20	18	16	14	12	10	8	146
1988	16	18	20	18	16	14	12	10	8	6	138
1989	18	20	18	16	14	12	10	8	6	4	126
1990	20	18	16	14	12	10	8	6	4	2	110
1991	18	16	14	12	10	8	6	4	2	0	90
1992	16	14	12	10	8	6	4	2	0	2	74
1993	14	12	10	8	6	4	2	0	2	4	62
1994	12	10	8	6	4	2	0	2	4	6	54
1995	10	8	6	4	2	0	2	4	6	8	50
1996	8	6	4	2	0	2	4	6	8	10	50
1997	6	4	2	0	2	4	6	8	10	12	54
1998	4	2	0	2	4	6	8	10	12	14	62
1999	2	0	2	4	6	8	10	12	14	16	74
2000	0	2	4	6	8	10	12	14	16	18	90
2001	2	4	6	8	10	10	14	16	18	20	110

Appendix Table A-6. Data for Figure 3, School 2

Year cohort completes grade 10	Average gain by grade										Average achievement in grade 10
	1	2	3	4	5	6	7	8	9	10	
1981	18	16	14	12	10	8	6	4	2	0	90
1982	16	14	12	10	8	6	4	2	0	2	74
1983	14	12	10	8	6	4	2	0	2	4	62
1984	12	10	8	6	4	2	0	2	4	6	54
1985	10	8	6	4	2	0	2	4	6	8	50
1986	8	6	4	2	0	2	4	6	8	10	50
1987	6	4	2	0	2	4	6	8	10	12	54
1988	4	2	0	2	4	6	8	10	12	14	62
1989	2	0	2	4	6	8	10	12	14	16	74
1990	0	2	4	6	8	10	12	14	16	18	90
1991	2	4	6	8	10	12	14	16	18	20	110
1992	4	6	8	10	12	14	16	18	20	18	126
1993	6	8	10	12	14	16	18	20	18	16	138
1994	8	10	12	14	16	18	20	18	16	14	146
1995	10	12	14	16	18	20	18	16	14	12	150
1996	12	14	16	18	20	18	16	14	12	10	150
1997	14	16	18	20	18	16	14	12	10	8	146
1998	16	18	20	18	16	14	12	10	8	6	138
1999	18	20	18	16	14	12	10	8	6	4	126
2000	20	18	16	14	12	10	8	6	4	2	110
2001	18	16	14	12	10	8	6	4	2	0	90

References

- Clune, William H. 1991. "Systemic Educational Policy." Wisconsin Center For Educational Policy, University of Wisconsin-Madison, July.
- Dominitz, Jeff, and Robert H. Meyer. 1991. "Educational Performance Indicators and School Report Cards: Lessons From Three States." University of Wisconsin-Madison. Mimeo.
- Dossey, John A., et al. 1988. *The Mathematics Report Card: Are We Measuring Up?* Princeton, N.J.: Educational Testing Services.
- Dyer, Henry S., Robert L. Linn, and Michael J. Patton. 1969. "A Comparison of Four Methods of Obtaining Discrepancy Measures Based on Observed and Predicted School System Means on Achievement Tests," *American Educational Research Journal* 6, 4 (November): 591–605.
- Goldman, Jay P. 1990. "Grading Schools Through Report Cards: Realtors, News Media Collect them For Comparative Purposes," *School Administrator* 47, 6: 26–30.
- Hanushek, Eric A. 1972. *Education and Race*. Lexington, MA: D.C. Heath.
- Hanushek, Eric A., and Lori Taylor. 1990. "Alternative Assessments of the Performance of Schools," *Journal of Human Resources* 25, 2 (Spring): 179–201.
- Henderson, James M., and Richard E. Quandt. 1971. *Microeconomic Theory: A Mathematical Approach*. New York: McGraw-Hill.
- Murnane, Richard J. 1975. *The Impact of School Resources on the Learning of Inner City Children*. Cambridge, MA: Ballinger.
- Powell, Brian, and Lala Carr Steelman. 1984. "Variations in State SAT Performance: Meaningful or Misleading," *Harvard Educational Review*, 54, no. 4: 389–412.
- Smith, Marshall S., and Jennifer O'Day. 1990. "Systemic School Reform." In *The Politics of Curriculum and Testing*. Edited by Susan Fuhrman and Betty Malen. 1990 Yearbook of the Politics and Education Association. London: Taylor and Francis. Pp. 233–67.
- U.S. Department of Education. 1971. *America 2000: An Education Strategy*.
- Wainer, Howard. 1986. "The SAT as a Social Indicator: A Pretty Bad Idea." In *Drawing Inferences from Self-Selected Samples*. Edited by Howard Wainer. New York: Springer-Verlag. Pp. 7–22.

College Choice, Academic Achievement and Future Earnings

Estelle James
State University of New York
and
World Bank

Nabeel Alsalam
National Center for Educational Statistics

A major issue in the literature on the economics of education is: Do schools make a difference? Do different schools produce different outcomes, and if so, why? Numerous studies have analyzed this question at the primary and secondary levels, usually using cognitive learning as the output variable. Only a small number of studies have tackled this issue at the college level, where the research interest has largely focused on the returns to quantity rather than quality. These studies have used future earnings or occupational status as the output variable, consistent with the premise that an important function of education is to improve one's position in the labor market. Colleges can enhance productivity and earnings by imparting general or specific skills, or information that helps students make good choices about their future career directions. Studies of college quality, like those of the quality of primary and secondary schools, have come up with ambiguous and contradictory results.

For example, in his analysis of the NBER Thorndike earnings data (for a group of World War II veterans whose earnings were measured in 1969), Wales (1973) found that graduates of top colleges (as measured by the Gourman rating, which is related to selectivity) received significantly higher earnings. Using the same data, Wachtel (1976) found that college expenditures per student exert a positive effect on earnings, and Solmon and Wachtel (1975) found that college type, as measured by Carnegie classification, also matters. According to Reed and Miller (1970), college rank (as measured by the average verbal and

mathematical aptitude of entering freshmen) had a positive effect on the weekly earnings of a sample of men surveyed by the Census Bureau in 1967. However, most of the students covered by these studies were in college prior to the vast expansion of the 1960s, which changed the nature of the higher education industry. In these regressions, usually only one college characteristic is specified as an indicator of quality. And only a small set of student background variables is included. Astin and others have argued that the impact of college quality is minimal once these are controlled (Astin 1968; Griffin and Alexander 1978). Indeed, a study by Alwin (1974) found only a small relationship between occupational success and college characteristics, after controlling for student composition.

One key problem encountered by all these studies has been the difficulty in obtaining detailed information about student characteristics, in order to control for ability, family influence, and prior education, which is essential if the purpose is to calculate the value added by the school. Longitudinal data tracking the individuals in a cohort are particularly scarce. Another problem has been the difficulty in obtaining disaggregated information about schools. A final problem is that we rarely have data about the student's college experience and academic achievement, such as choice of courses, major and GPA, that may influence earnings. (See Hanushek 1986, for a thorough analysis of these and other problems that beset the "school effectiveness" and "educational productivity" literature.)

In this paper we exploit a uniquely rich data set, the National Longitudinal Study of the High School Class of 1972 (NLS-72) and the Post-secondary Education Transcript Study (PETS), to overcome these problems and answer the questions: Does it matter which college a student attends? If it matters, which college characteristics lead to higher earnings? Do higher expenditures or a more selective student body imply superior results? Which is better, a large research university or a small liberal arts college? Does the public/private typology make a difference, as some feel it does at the secondary level? We also examine the impact on earnings of student behavior while at college. Are academic achievement and curriculum choice harbingers of future achievement? (For a preliminary analysis of these issues see James, Alsalam, Conaty, and To 1989.)

We begin by setting forth our methodology, including our model, data sources, choice of variables, and statistical technique. In the second part of this paper we present our findings concerning the effects of college characteristics, and in the third we give our results concerning the effects of other aspects of the college experience, particularly choice of major. The fourth section is a summary of our results and their limitations.

Briefly, our fixed-effects model shows that the particular college attended does indeed make a difference, explaining 17 to 29 percent of the variance in earnings. However, we are unable to tie this college effect down to observable college characteristics, which taken as a group explain only 1 to 2 percent of the variance in earnings. These effects exist, but they are very small.

Moreover, whether fixed effects or observable characteristics are measured, the college effect becomes statistically insignificant once family background, labor market experience, and major are controlled. To a large extent the world perceives a differential college effect because it is perceiving "gross output" rather than "value-added" and is not taking account of the many other factors that affect earnings, some of which are correlated with choice of college.

In contrast to these negative findings about college effects, we find that what a student does while in college strongly affects future earnings, even after all the other variables in our model are controlled. Apparently, direct measures of skill acquisition matter more than indirect screening by college characteristics.

Methodology

The Model

As is well known, there are multiple outputs of higher education, including knowledge gained, earning power enhanced, values formed, amenities consumed, and research undertaken, all of which enter into student and social utility functions. In this paper we concentrate on one output of education—future earnings—as a function of all the jointly supplied inputs.

We model earnings (Y) as a function of four sets of variables:

- X_1 = a set of individual characteristics including family background and prior academic achievement;
- X_2 = a set of institutional characteristics, including college expenditures, college organization, and student body composition, which determine the value added by the college;
- X_3 = higher educational experience variables that are chosen by the student but may be influenced by the college;
- X_4 = labor market variables, such as experience and weeks worked per year.

Our focus here is on the impact on earnings of institutional characteristics and other aspects of the higher educational experience.

Before presenting our results we outline the data and expected direction of causal relationships, and discuss our treatment of several methodological problems.

Data Sources and Sample

Information about student characteristics, earnings, and other labor market variables comes from the National Longitudinal Study of the High School Senior Class of 1972, which follows this cohort through further education and into the labor market. This survey gives detailed information about family background, education, and academic achievement prior to entering college, as well as subsequent labor market experience. The fifth follow-up in 1986 includes 12,841 men and women. Two-thirds of the total had some postsecondary education and one-fourth had received their college degree. We deal in this paper with a subset of the latter group, the 1,321 males whose graduating institution was identified, who took at least sixty credit hours in that institution, and who worked for an employer at least twenty hours per week in 1985. Most of them had been out of college for seven to nine years.

There are 499 colleges and universities in our subsample. Over half are Ph.D.-granting institutions, three-quarters are public, and enrollment size varies from 288 to 50,011. Many of the smaller colleges enrolled only one member of our sample, but 285 institutions (typi-

cally the larger ones) had two or more observations, totaling 1,107 male students; we used these as a subsample for the fixed-effects models. We obtained most of our data about the characteristics of these colleges and universities from the Higher Education General Information Survey (HEGIS), which conducts annual surveys of postsecondary four-year institutions; we chose 1975 as a representative year for our cohort. This was supplemented by data from Cass and Birnbaum (1975).

Finally, the Postsecondary Education Transcript Study gave us the college transcripts of each student, from which we derived the college experience variables. None of the previous studies on college quality has had access to such detailed information about curriculum choices and achievement in college. All financial data were inflated to 1986 prices using the Consumer Price Index.

Institutional Characteristics

We view the college consumer as purchasing a set of characteristics that is experienced uniformly by all students at a given institution. In our fixed-effects model we identify each college by a dummy variable that captures all its observable plus unobservable characteristics. In our ordinary least squares (OLS) models we examine the observable characteristics in greater detail. This section describes some of the observable variables we use.

The school effectiveness literature has been particularly interested in the influence of "expenditures per student." If colleges were competitive, privately financed, profit-maximizing institutions, and if consumers cared primarily about future earnings and had full information about educational production functions, higher costs would have to be covered by higher prices which would be sustainable only if they led to higher future earnings. However, the institutions in our sample are public or nonprofit, much of their revenues coming from state legislators or donors, and devices such as accreditation procedures and reputation limit entry. Under these circumstances, colleges may have potential profits with which to pursue their own discretionary goals, their nonconsumer funders may support multiple objectives, and students may be uninformed about college effects or may care about many outputs of education besides future productivity. As a result of all

these factors, we do not have a strong *a priori* reason to expect a systematic relationship between college expenditures and future earnings.

To investigate the expenditure effect, we started with the idea of decomposing college spending into its separate components—expenditures on instruction, research, institutional support services, and financial aid. However, strong multi-collinearity precluded this strategy. Therefore we used, alternatively, the most inclusive measure, logged educational and general spending per student (LXPS); the most directly relevant measure, logged instructional expenditures per student (LINSXPS); and a combination of inputs—the student-faculty ratio (S/F) and average faculty salaries (FACSAL). Ideally, we would have adjusted our measures of monetary inputs for regional cost differences, but unfortunately we did not have access to the detailed city-by-city, education-relevant cost-of-living index that would be necessary to convert these monetary inputs into real inputs. However, our experiments with S/F captured the basic idea. As it turned out, the results for all the cost variables were very similar, and those with LXPS are presented in this paper.

We wanted to investigate whether institutions under public versus private control, with research versus teaching missions, and with predominantly graduate versus undergraduate student bodies, behave differently. Theory suggests that such differences exist, but the direction and magnitude are ambiguous. For example, private institutions may utilize their resources more efficiently than public ones and may benefit from a halo effect in the labor market or on the other hand may have to devote substantial resources to fundraising (see James 1989). The presence of doctoral students may lead to a diversion of resources away from undergraduates or conversely may add a pool of cheap labor available to teach undergraduates. And similarly, research may enhance or detract from the undergraduate teaching function (see James 1978; James and Neuberger 1981). These possible effects were captured by entering private colleges, Ph.D.-awarding institutions, and Research Type I Universities (Carnegie classification) as dummies (PRIV, PRIV*E, PHD, RES) or, alternatively, as interactions with expenditures per student, and by including the percent of graduate students in total enrollment (PGRAD). The total number of FTE students (LFTE) was entered to allow for economies or diseconomies of scale.

Regional dummies were included to capture regional fixed effects. Only East (E) had any impact and was left in the final regression.

We were particularly interested in the impact of the average SAT score of entering freshmen as an index of institutional selectivity (SEL). Selectivity may influence earnings in several ways: it may raise the amount of learning, hence the acquisition of general human capital at the institution through the peer group effect; it may be an informational signal to employers about the probable aptitude of individual students (a version of the screening hypothesis); and students from a more selective institution may benefit from its prestige or from the social network that it generates.

Higher Educational Experience

While in college the student makes a number of choices—concerning major, curriculum, how hard to work to obtain a high GPA, and whether to proceed to a postgraduate degree. How do these choices influence future earnings? We examine the impact of choosing majors in Engineering (ENG), Business (BUS), Humanities and Fine Arts (HUM), Social Science (SOCSCI), Math plus Physical Science (M&SCI), Biological Science (BIO), Health Professions (HEALTH), and all others (OTHER), relative to education, the omitted category. The choice of major influences the range of occupational options that will be open to a student later in life; therefore we would expect some of these dummies to be positive and others to be negative. We also included the number of college math credits and Math GPA; our final specification uses the interaction of these two terms (MATH*GPA). Finally, we introduce the college GPA as an independent variable, on grounds that this may signal cognitive skills and traits such as discipline and perseverance, to the employer. Postgraduate degrees are broken down into MA and all others (HIDEG); the latter, in particular, is expected to have a positive effect.

Background Characteristics

A voluminous literature documents the fact that family background and prior academic achievement strongly influence future earnings, much of this operating through their impact on the choice of college

and probability of completing college. We would expect background effects on earnings to be much smaller in this study, where everyone has a college degree and characteristics of the college are explicitly included. However, some background variables may still be important, in part because they serve as proxies for unobserved productivity-enhancing student characteristics such as ambition, learning acquired in the household, and access to labor market information. Including these variables should therefore increase the explanatory power of our model and reduce the potential bias in the estimated coefficients of the college variables with which they may be correlated. Student background variables included as controls were race or nationality (BLACK, BLK-SOUTH, HISP), religion (CATH, JEW), parental income (FAMINC), father's education (FAED), attendance at Catholic high school (CATHHS) and other private high schools (OTHPVT), percentile rank in high school (PRANK), a dummy for playing a leadership role in high school newspaper, student government, or athletics (XCURR), and the individual's verbal plus mathematics SAT score or a transformation of ACT scores into SAT (SAT). (For the method used to convert ACT to SAT, see Astin 1971.)

Labor Market Experience

While our focus is on undergraduate education, in some equations we control for a host of basic market variables that influence earnings: total months of employment since degree (EXP), tenure on current job (TEN), weeks worked in 1985 (LWW), hours worked per week (LHW, which we treat as a two-part spline allowing for the possibility that returns per hour may differ for hours worked under or over 35 per week), and a dummy for career interruptions that exceeded one year (INTERR). Marital status and number of children in 1985 were included as variables that might influence unobservable labor market choices such as effort.

Specification Problems and Alternative Models

A problem that always arises in the school effects literature concerns biases that may be introduced by unobserved elements of student

and institutional characteristics, and by the endogeneity of elements of some college and labor market variables. For example:

1. The omission of some institutional characteristics may understate the total college effect. We deal with this problem by including a dummy variable for each college, which captures its full observable and unobservable characteristics in the fixed-effects model.

2. The correlation between unobserved student characteristics and observed institutional characteristics may bias the college effect. Suppose that, as a result of student choice and college admissions procedures, ambitious, hard working students end up in colleges with a high selectivity index and expenditure rate. These colleges would appear to increase the earnings of their students, whereas actually they were simply choosing students who would have earned a lot anyway due to their ambition. We have tried to minimize this problem by including numerous observable student variables (such as high school rank and extra-curricular activities), and college experience variables (such as curriculum and GPA) that may be correlated with and therefore proxy the unobserved characteristics. In some equations, we used instrumental variables for SEL and LXPS.

3. Some colleges may have policies that induce their students to enter remunerative majors and to acquire postgraduate degrees (for example, these colleges may not offer low-paid education majors, may require math, and may help their students gain entry to graduate and professional schools). If so, controlling for the higher educational variables understates the college effect. We deal with this problem, in part, by running our regressions with and without the higher educational variables. If college policies determine curriculum choice, the college effect on earnings should decline when curriculum is controlled.

4. The effect of the higher educational experience variables may also be biased by their correlation with unobserved student characteristics; e.g., students who choose to take math courses may be smart and hard working and hence earn high wages because of these characteristics, not because of the skills they acquired while taking math. We considered using two-stage techniques here, but it is difficult to find instruments that are clearly exogenous, can confidently be excluded from the wage equation, and are good predictors of curriculum choice. Thus, we have not been able to eliminate the possibility that the large

effects found for the college experience variables are due, in part, to their correlation with unobserved student attributes.

5. Similar specification problems could be outlined with respect to the labor market variables.

Despite all these limitations, we believe that a clear picture does emerge of the impact of college characteristics and college experience variables, since our basic results are robust with regard to the various specifications explored.

College Effects

Tables 1 and 2 present our estimate of logged annual earnings for 1985 (LY 85) as a function of college characteristics (X_2) and higher educational experience (X_3), in some cases controlling for student's background (X_1), which precedes college, and labor market choices (X_4), which follow college. Lifetime earnings would obviously have been a preferable but unavailable dependent variable. Our indicator of earnings seven to nine years after college is a much better indicator than starting wages, which have sometimes been used in studying school effects.

We present a series of equations that starts with X_2 and sequentially adds X_1 , X_3 , and X_4 . Table 1 uses fixed college effects and is based on a subsample of 1,107 students in 285 colleges with two or more students. Table 2 uses OLS with observable college characteristics in place of the institutional dummies and is based on the full sample of 1321 students and 499 colleges. In addition to giving the individual variables, we also calculate the proportion of variance explained by the individual, institutional, higher educational experience, and labor market variables as a group. When all the variables are included, we are able to explain over half the variance in earnings of this cohort of college graduates.

Fixed Effects

Column (1) of table 1 presents a "gross output" fixed-effects model that replicates "what the world sees." Each college is represented by a

Table 1. 1985 Annual Earning Regressions for Males with Fixed College Effects Explanatory Power of Groups of Variables (2+ sample)

	College dummies (1)	College + background (2)	College + background + LM (3)	College + background + LM + higher ed (4)
R^2				
Entire model	.333 ^a	.365 ^a	.507 ^a	.56 ^a
College dummies	.333 ^a	.286 ^b	.211 ^c	.171
Background		.032 ^a	.022 ^b	.015 ^c
Labor market			.142 ^a	.124 ^a
Higher ed. exp.				.053 ^a

NOTES: Each regression is estimated for a sample of 1,107 males from the High School Class of 1972 who eventually received a bachelor's degree, from 285 colleges, each of which had two or more students in our sample.

a. Significant at .1% level

b. Significant at 1% level

c. Significant at 5% level

Background variables that were significant in at least one equation in Tables 1 or 2 are family income, SAT score, Black, Black*South interaction, Catholic, Catholic H.S., other private high school. See text for other background variables that were included.

Labor Market variables are experience, tenure on current job, log weeks worked last year, hours worked per week, marital status, number children, career interruption.

Higher Education variables are dummies for different majors (Business, Engineering, Math-Science, Humanities, Social Science, Biology, Health Science, Other, Education as omitted category), GPA, Math*GPA, MA degree, higher degree.

Table 2. Annual Earnings Regressions for Males with Observable College Characteristics (full sample)

	College characteristics (1)	College + background + LM (2)	College + background + LM (3)	College + background + higher ed (4)
College characteristics				
SEL	.051 (2.93) ^b	.026 (1.42)	.017 (1.06)	.012 (.8)
LFTE	.08 (3.48) ^a	.064 (2.74) ^b	.048 (2.32) ^c	.037 (1.84) ^d
PRIV	.004 (.08)	-.029 (.54)	-.05 (1.04)	-.005 (.11)
EAST	-.025 (.55)	-.04 (.9)	-.052 (1.3)	-.008 (.20)
PRIV*E	.149 (2.1) ^c	.136 (1.94) ^c	.182 (2.93) ^b	.106 (1.76) ^d
PGRAD	-.001 (.5)	-.002 (.81)	-.001 (.59)	-.001 (.38)
LXPS	.003 (.05)	.011 (.21)	.035 (.76)	-.009 (.21)
LXPS*RES	-.003 (.62)	-.005 (.94)	-.003 (.6)	-.003 (.74)
Higher education experience				
BUSMAJ				.326 (6.89) ^a
ENGMAJ				.484 (7.75) ^a
M&SCI				.273 (4.05) ^a
HUM				.051 (.82)
SOCSCI				.259 (5.32) ^a

Table 2 (continued). Annual Earnings Regressions for Males with Observable College Characteristics (full sample)

	College characteristics (1)	College + background + LM (2)	College + background + LM (3)	College + background + higher ed (4)
BIO				.238 (3.78) ^a
HEALTH				.243 (3.35) ^a
OTHER				.189 (3.83) ^a
GPA				.066 (2.0) ^c
MATH*GPA				.002 (3.45) ^a
MA				.05 (1.45)
HIDEG				.151 (2.59) ^b
Constant	9.013 (20.96) ^a	8.984 (20.8) ^a	-1.018 (.83)	-.67 (.22)
R^2				
Entire model	.038 ^a	.09 ^a	.301 ^a	.382 ^a
College characteristics	.038 ^a	.014 ^c	.013 ^b	.005
Background		.039 ^a	.03 ^b	.025 ^b
Labor market			.211 ^a	.202 ^a
Higher ed. exp.				.079 ^a

NOTES: Each regression is estimated for a sample of 1321 males from the High School Class of 1972 who eventually received a bachelor's degree from 499 colleges. See Table 1 for significance levels and other variables that were included in regressions.

a. Significant at .1% level

b. Significant at 1% level

c. Significant at 5% level

d. Significant at 10% level.

dummy that captures its fixed effect. Consistent with conventional wisdom, the college that a student attends makes a significant difference in this model, explaining 33 percent of the variance. This specification maximizes the college effect, since it subsumes both observable and unobservable college characteristics (including 285 dummy variables) and does not control for anything else.

Of course, much of this effect stems from the background of the students, which is not randomly distributed across colleges. Thus, in column (2) we include both institutional dummies and student background, in an effort to measure "value added" by the college. Although the proportion of variance explained by the college dummies declines (to 29 percent), their effect remains highly significant. In column (3) we enter the basic labor market variables, and column (4) adds the higher educational experience variables. As discussed above, these explanatory variables may in part be capturing college policies or correlated student characteristics that determine earnings. Therefore, their inclusion should increase the explanatory power of our model and at the same time decrease the apparent college effects. Indeed this is what happens. In the last equation, the fixed college effects as a group explain only 16 to 17 percent of the variance in earnings, half of the original amount, and given the large number of degrees of freedom used up, this is no longer significant. For reasons given earlier, we believe the "true" size of the effect is smaller than that in equation (1), but larger than that in equation (4); i.e., it explains between 17 and 29 percent of the variance in earnings and is marginally significant.

Observable College Characteristics

Table 2 replicates these results using observable college characteristics in place of the separate dummy variables for each institution. These equations attempt to ascertain where the college effect is coming from. We are not very successful in that regard, since the observed characteristics as a group explain only 1 to 2 percent of the variance in earnings, once other variables are controlled.

Column (1) includes college characteristics only, and column (2) adds family background. Column (3) is closest in spirit to earlier estimations by Wachtel (1976), Solmon and Wachtel (1975), and Reed and Miller (1970), and our results are similar (positive but small college

effects)—increasing our confidence in these findings. The differences in our overall conclusions, then, stem from the richer set of variables we have been able to use. When both higher educational experience and labor market variables are included, as in column (4), neither the individual college characteristics nor their sum has a significant effect.

To the extent that observable college variables matter, it appears that characteristics associated with size, control, and possibly student body composition matter most. A doubling in enrollments increases earnings 4 to 6 percent, an indication of economies of scale in producing future earnings that may stem from greater program variety (enabling better student-major matches) or from greater visibility and access to labor market information. We obtain ambiguous results on the advantages of attending a private college, consistent with our ambiguous predictions: this has a positive effect (of 10 to 13 percent) in the East (where private colleges have long-standing labor market connections), but not in the rest of the country. As expected, SEL has a positive effect in columns (1)–(3)—a 100-point increase in SAT of freshman class increases earnings 3 to 5 percent—but its size declines and significance disappears when higher educational experience and labor market variables are in the equation, as in column (4).

In contrast to the college variables that matter, expenditure per student (LXPS) has a small coefficient that is never close to significance. This contradicts Wachtel's earlier finding but is consistent with Morgan and Duncan's (1979) and with much of the literature on primary and secondary school effectiveness.

We thought that the high-spending institutions might be universities that allocate much of their resources to graduate programs and research. To test this possibility, we tried specifications in which LXPS was interacted with Ph.D.-granting or Research Type I universities and/or dummies added for Ph.D.-granting institutions, research institutions, and percentage of graduate students. The interaction terms were always negative, and the main LXPS effect became more positive (or less negative), but nothing we did ever made it significant. Apparently research and graduate programs do not increase or decrease future undergraduate earnings, and they also do not explain the unimportance of LXPS. We obtained similar results when instructional expenses, faculty salaries, and S/F ratios were used instead of LXPS. Based on the

experience of this sample, it appears that attending a higher-spending college is not the way to increase future earnings.

Alternative Specifications

Because these negative findings about selectivity and expenditures per student run counter to the conventional wisdom, we explored alternative specifications to see whether and under what conditions positive effects might emerge (see table 3).

1. We attempted to deal with the endogeneity problem by using instrumental variables to predict the selectivity of a student's college; the predicted selectivity value was then used in place of the actual value in column (1). We did the same for expenditures per student. The coefficients on SEL and LXPS rise when instrumental variables are used, but neither equation changes our basic conclusion that LXPS is always insignificant, and SEL is insignificant when higher educational experience and labor market variables are in the equation.

2. We considered the possibility that multi-collinearity among college characteristics was hiding the true significant effect of SEL, so we omitted all other institutional variables from the equation. We did the same for LXPS. As expected, the size of the SEL and LXPS coefficients increased in these specifications, but they remained insignificant when higher educational experience and labor market variables were both in the equation; and, of course, the total college effect declined.

3. Finally, we interacted SEL (and LXPS) with several student, college, higher education, and occupational variables to ascertain whether our uniform-effects model is understating the true effect (e.g., see Summers and Wolfe 1977). Perhaps SEL (or LXPS) has a higher payoff to students with higher SAT scores, or for those entering science majors, professional or managerial occupations, or going on to higher degrees. In general these interaction terms were insignificant. We conclude that if an interactive model is appropriate, these data are simply not strong enough to detect it.

Through all these specifications, our best estimate remains that the college a student chooses does make a marginal difference, that this difference becomes insignificant as additional explanatory variables are added to the model, and that only a very small fraction can be ascribed to college characteristics readily observed and measured.

Table 3. Alternative Specifications with SEL LXPS, and X₃

	SEL (1)	LXPS (2)	SEL only (3)	LXPS only (4)
<i>R</i> ²	.383 ^a	.382 ^a	.378 ^a	.377 ^a
SEL	.069 (1.48)	.011 (.75)	.017 (1.48)	-
LFTE	.037 (1.88) ^d	.037 (1.83) ^d	-	-
PRIV	-.007 (.17)	-.007 (.16)	-	-
E	-.049 (.97)	-.009 (.24)	-	-
PRIV*E	.116 (1.91) ^d	.109 (1.79) ^d	-	
PGRAD	-.001 (.25)	-.001 (.43)	-	-
LXPS	.002 (.06)	.043 (.28)	-	.022 (.59)
LXPS*RES	-.004 (.84)	-.004 (.96)	-	-.001 (.21)

Notes: Columns (1) and (2) use instrumental variables to predict SEL and LXPS, respectively, and then use the predicted values in place of the actual values. The full equation includes College Characteristics, Background, Labor Market, and Higher Education Experience, as in equation (4), Tables 1 and 2. Columns 3 and 4 omit all college characteristics except SEL and LXPS, respectively; full equation includes Background, Labor Market and Higher Education Experience.

a. Significant at .1% level.

b. Significant at 1% level.

c. Significant at 5% level

d. Significant at 10% level.

Table 4. Means and Standard Deviations of Variables

	Mean	Standard deviation
LY85	10.26	.52
Institutional variables		
SEL	9.87	1.16
LFTE	9.031	1.046
PRIV	0.248	
EAST	0.202	
PRIV*E	0.091	
PGRAD	12.267	8.800
LXPS	8.840	0.430
LXPS*RES	1.630	3.558
Higher educational experience		
MATH*GPA	23.785	33.444
GPA	2.914	0.464
BUSMAJ	0.230	
ENGMAJ	0.103	
M&SCI	0.058	
HUM	0.061	
SOCSCI	0.178	
BIO	0.062	
HEALTH	0.027	
OTHER	0.166	
MA	0.140	
HIDEG	0.054	

The Higher Educational Experience

Explanatory Power

While institutional characteristics do not explain a large proportion of the variance in earnings, other aspects of the higher educational experience, such as choice of major, number of math credits taken, GPA, and postgraduate degree matter a great deal. All of these variables are highly significant, add substantially to the R^2 of the model and, as a group, explain 3 to 8 percent of the variance in earnings in the OLS regressions, more than the observable college and student characteristics put together. They explain somewhat less—2 to 5 percent—in the fixed-effects model, consistent with the hypothesis that the higher education variables are determined, in part, by unobserved college policies. (See tables 1 and 2.)

Students with a higher GPA have higher expected earnings; when GPA increases from *C* to *B* or from *B* to *A*, annual earnings rise 7 to 8 percent. (Also see Wise 1975.) While GPA indicates, in part, that students have acquired specific knowledge, we prefer to think of it as a proxy for unobservable characteristics such as ability combined with inputs of time and effort, general human capital characteristics that also lead to higher productivity in the labor market. This interpretation is supported by the behavior of the student's SAT, which has a positive effect in equations (2) and (3) that becomes negative when GPA enters in equation (4). Suppose that the SAT is a proxy for academic ability or potential achievement, which must be combined with inputs of time and effort to produce actual achievement. If the combination is present it produces achievement in college and thereafter; but if the input of effort is not forthcoming in college, it probably will not be forthcoming at work either. By this interpretation, ability alone does not generate significantly higher grades or earnings; the payoff to GPA is a payoff to the combination of ability and effort.

The positive return to college math is also noteworthy for educational policy. Taking three additional math credits (usually one course) and receiving an *A* increases earnings 1 to 2 percent. This is not surprising since math governs entry to certain highly-paid occupations

such as engineering. (For a further discussion of the impact of math, see Alsalam 1989a.)

A higher degree, particularly the Ph.D., LL.D or M.D., also has a significantly positive effect, raising earnings 10 to 15 percent (and probably more at a later point in the age-earnings life cycle).

This leads us to the important issue of returns to major. Which majors are worth more in the marketplace—those imparting general or specific skills? While there are large differences in the returns to different majors, in general we cannot say that either vocational or liberal arts majors have an advantage. For example, Education (which is the omitted category) is one of the least remunerative majors, but Engineering and Business, also vocational majors, are on top, 30 to 40 percent higher than Education. The Physical and Social Sciences are in between, 20 to 30 percent higher than the other liberal arts subjects, Humanities and Fine Arts.

Equilibrium Wage Differentials by Major and Social Efficiency

How are these differences in returns to majors sustainable, and would society be better off if students were induced to move to the higher-paying (and presumably higher-productivity) majors? The answers to these two questions are interrelated.

One well-known explanation for why this can be an equilibrium situation is that different majors lead to jobs with different combinations of pecuniary and nonpecuniary benefits, and different students have different preferences between these and hence make different choices. (For example, this might explain the low return to Humanities majors and the high return to Business majors who become managers.) In this case the differential returns to majors reflect differential tastes, are sustainable in the long run, and as long as students have accurate information, it would not be efficient to shift more of them involuntarily into higher-paying majors.

A second explanation is that students differ in innate ability and/or work effort, that these differences lead them to choose “hard” or “easy” majors and to earn high or low wages. (For example, the entry and exit requirements are probably more demanding for Engineering majors than Education majors.) As a variant on this explanation, some majors may have “gatekeeping” courses, such as math, which some

people find hard, others easy; those with math aptitude are more likely to choose these majors and to earn a rent on their scarce aptitude. According to this interpretation, we cannot assume that people who choose different courses and majors are otherwise similar in ability or effort or attribute their earnings differential to their choice of major or to specific skills they have acquired in college; furthermore, if more students entered higher-paying majors, they may well have lower productivity and earnings than current students.

A third explanation focuses on the reasons why some majors are “hard” and others “easy” and interprets these differences as an institutional response to situations where large differences in the real cost of training students exist across majors, and “society” believes that ability or effort, rather than price, should be used to ration space in high-cost fields in order to avoid myopic choices of majors and jobs. If price is the same for all programs of study (as tends to be the case at the undergraduate level), but barriers to entry and exit vary, monetary returns will also vary; and this may be both sustainable and efficient. Under this scenario, we would expect majors with the highest training costs to have the greatest restrictions and hence the highest private monetary returns. Impressionistically, we seem to observe this relationship at the extremes (for example, this may explain the entry and exit difference between Engineering and Education), but there does not seem to be a good fit among fields in the middle.

Conclusion

In conclusion, we have defined college quality as a multidimensional concept, and this paper has concentrated on one dimension—value added to future earnings, a proxy for future labor market productivity. Our rich longitudinal data source has allowed us to control for an unusually wide range of incoming student characteristics, thereby yielding better estimates than previously possible of the impact of college and curriculum choice on future earnings.

Briefly, our fixed-effects model indicates that a student’s choice of college does make a difference, explaining 17 to 29 percent of the variance in earnings. However, we are unable, using OLS, to tie this col-

lege effect down to observable college characteristics, which, taken as a group, explain only 1 to 2 percent of the variance. Thus, prospective students and the families who make their decisions based on these readily accessible indicators will not be getting much pecuniary return; on the other hand, investing time and effort to collect information about the "unobservable" characteristics of colleges may pay off.

The observable characteristics, as well as the total college fixed effect, become insignificant once curriculum choice and labor market experience are included in the equation. While controlling for all these variables may lead us to underestimate the total college contribution, omitted student variables may have the opposite effect. If the former bias exceeds the latter, we may conclude that the true size of the college effect is greater than that in column (4) of table 1 but less than that in column (1) and is marginally significant. Experiments with random-effects models and instrumental variables did not change this basic conclusion.

In particular, expenditures per student never have a large or significant effect, and college selectivity, which is widely believed to predict success, has effects that are both small and insignificant once other variables are in the equation. Findings concerning college selectivity cast some doubt on the screening hypothesis, since this is a logical candidate for employers to use as a proxy for ability. According to these results, they do not do so.

In contrast to the limited effects of college choice, what a student does while in college (which is presumably related to the human capital acquired there), strongly affects future earnings, even after all other variables in our model are controlled. Grades, major, math courses, and further degree are all highly significant and, taken as a group, explain 3 to 8 percent of the variance in earnings, more than measured college and student characteristics put together. This finding casts further doubt on the screening hypothesis.

Several caveats are in order at this point. First of all, part of the large effect of the higher educational experience variables probably stems from the unobservable productivity-enhancing student and college characteristics with which they are correlated, but which we could not capture. Second, the relative returns to different majors and occupations may change through time as a function of supply and demand, so it is impossible to generalize from this cohort to all cohorts. Third, the

relative earnings of different majors and occupations may change over the lifetime of this cohort as a function of their training and its returns. For example, the doctors and lawyers in our sample had relatively little experience and are probably at the start of a steeply rising age-earnings profile, while the opposite is true for nurses and teachers.

Fourth, we have assumed uniform college effects; if a thoroughly interactive model is more appropriate, we may have underestimated college effects. (This sample is not suitable for analyzing such a model.) Finally, these results apply to men only; a paper on women is now in process, and the picture appears quite different. But at this point, we would have to conclude that if quality and output are defined in terms of future earnings and productivity, high inputs do not necessarily lead to high outputs, and indeed there is no easy way to identify high-quality colleges.

Can it be inferred that students who spend long hours trying to get accepted into selective colleges, that parents who pay high tuition, and that colleges with large expenditures are all wasting their time and money? Not necessarily. As we said at the beginning, colleges produce many outputs, and higher future earnings is but one of them. Institutions may be interested in research as well as teaching, parents may be interested in the cognitive development and/or value formation of their children, students may be interested in the social ambience of their college, and all of these may be important to society at large. Indeed, much of the expenditures of higher-spending institutions may be directed toward these other ends, some of which may have a diffuse long-run impact on productivity. Therefore, it is imperative to undertake other studies with alternative output measures to get a complete picture of the determinants and consequences of college quality.

NOTES

1. There may be substantial multi-collinearity among these variables. We did not consider this a big problem since we are not trying in the paper to identify the separate effects of each student background variable, but rather to control them to avoid overestimating the value added by the college. We also included an *F* test for the joint significance of the group as a whole.

2. Switching to OLS in table 2 allows us to use the entire sample of colleges and students as well as to weight student differentially in order to adjust for the stratified sample design used by NLS and make them representative for the sample as a whole. Using the unweighted subsample to replicate table 2 indicates that the coefficients and significance of our variable are largely unchanged by the weighting procedure. We also used a random-effects or generalized least-

squares model, to take account of the fact that all the people who attend the same school are subject to a common component and therefore have less independent variation than OLS would presume. It turned out that the variance component due to the unobserved college effect is negligible; the size and significance levels of some variables increase, but most are unchanged. Therefore we present the fixed effects and OLS results in this paper.

3. Economies of scale were also found in Cohn, Rhine, and Santos 1989, who used student enrollments and research as outputs; i.e., average costs fall as a composite of these outputs rises.

List of Symbols and Data Sources

College Characteristics

SEL	Mean SAT Score of Incoming Freshman Class 1976.
LFTE	The log of the full-time equivalent enrollment.
PRIV	Dummy variable indicating private institutional control.
EAST	Dummy variable indicating attendance at a Northeastern college.
PRIV*E	Dummy variable indicating attendance at a Private Northeastern college.
PGRAD	Proportion of full-time-equivalent students who are graduate students.
LXPS	The log of total educational and general expenditures per full-time equivalent student, including expenditures on instruction, research, public services, libraries, academic support, student services, institutional support, operation and maintenance of plant, scholarships and fellowships, and educational and general mandatory transfers.
LXPS*RES	Interaction of LXPS and the Carnegie Classification of Type 1 institutions. Carnegie Institution "A Classification of Institutions of Higher Education." 1988 p. 1.

Higher Educational Experience

MATH*GPA	Total number of math, statistics, and computer science credits earned in college multiplied by the grade point average the student earned in these courses. Math credits are calculus level and above; i.e., precollegiate and collegiate math are excluded.
GPA	Grade point average at first undergraduate degree-granting institution.
BUSMAJ	Dummy variable indicating business major.
ENGMAJ	Dummy variable indicating engineering major.

M&SCI	Dummy variable indicating a physical science or mathematics major.
HUM	Dummy variable indicating humanities or fine arts major.
SOCSCI	Dummy variable indicating social science major.
BIO	Dummy variable indicating a life science major.
HEALTH	Dummy variable indicating a health science major.
OTHER	Dummy variable for all other major except education which is the omitted category.
MA	Dummy variable indicating receipt of a Master's degree.
HIDEG	Dummy variable indicating the receipt of a Ph.D. or professional degree above the master's level, such as law or medicine.

SOURCES: Riccobono, J., L. B. Henderson, G. J. Burkheimer, C. Place, J. R. Levinsohn *National Longitudinal Study: Base Year (1972) through Fourth Follow-Up (1979) Data File Users Manual*. Vols. 1-4 C. D. Carroll, et al., *The National Longitudinal Study of the High School Class of 1972 (NLS-72) Fifth Follow-Up (1986) Data File User's Manual*. Higher Education General Information Survey (HEGIS), 1975. Tuition and selectivity taken from Cass, James and Max Birnbaum, *Counselors' Comparative Guide to American Colleges: 1976 Edition* (New York: Harper and Row, 1975). Jones, C., R. Baker, and R. Borchers, *National Longitudinal Study of the High School Class of 1972: Postsecondary Education Transcript Study Data File User's Manual*, National Center for Educational Statistics, August 1986.

References

- Alsalam, Nabeel. 1989a. "College Quality and Future Earnings: More Math Means More Money." National Center for Educational Statistics, Mimeo.
- _____. 1989b. "Job Training for the Work-Bound Graduates of the High School Class of 1972: A Description with Contrasts and Consequences." National Center for Educational Statistics, Mimeo.
- Alwin, Duane. 1974. "College Effects on Educational and Occupational Attainments," *American Sociological Review* 39: 210–223.
- Astin, Alexander. 1968. "Undergraduate Achievement and Institutional Excellence," *Science* 161: 661–668.
- _____. 1971. *Predicting Academic Performance in College: Selectivity Data for 2300 American Colleges*. New York: The Free Press.
- Bryk, A. S., and Stephen W. Raudenbush. 1988. "Toward a More Appropriate Conceptualization of Research on School Effects: A Three-Level Hierarchical Linear Model," *American Journal of Education* 97, 1: 65–108.
- Cass, James, and Max Birnbaum. 1975. *Counselors' Comparative Guide to American Colleges*, 1976 edition. New York: Harper and Row.
- Cohn, Elchanan, Sherie Rhine, and Maria Santos. 1989. "Foregone Earnings of College Students in the U.S., 1970 and 1979: A Microanalytic Approach," *Higher Education* 18, 6: 681–95.
- Griffin, Larry J., and Karl L. Alexander. 1978. "Schooling and Socio-Economic Attainment: High School and College Influences," *American Journal of Sociology* 84, 2: 319–347.
- Hanushek, Eric A. 1986. "The Economics of Schooling: Production and Efficiency in Public Schools," *Journal of Economic Literature* 24: 1141–1177.
- James, Estelle. 1978. "Product Mix and Cost Disaggregation: A Reinterpretation of the Economics of Higher Education," *Journal of Human Resources* 13: 157–186.
- _____. 1989. "Decision Processes and Priorities in Higher Education." In *The Economics, Management and Fiscal Environment of American Universities*. Edited by S. Honeack. Albany: SUNY Press.
- James, Estelle, Nabeel Alsalam, Joseph Conaty, and Duc-Le To. 1989. "College Quality and Future Earnings: Where Should You Send Your Children to College?" *American Economic Review* 79: 247–252.
- James, Estelle, and Egon Neuberger. 1981. "The University Department as a Nonprofit Labor Cooperative," *Public Choice* 36: 585–612.
- Manski, Charles F., and David A. Wise. 1983. *College Choice in America*. Cambridge: Harvard University Press.

- Murnane, Richard J. 1986. "Comparisons of Public and Private Schools: What Can We Learn?" In *Private Education: Studies in Choice and Public Policy*, edited by Daniel Levy. New York: Oxford University Press, pp. 153–169.
- Morgan, James N., and Greg J. Duncan. 1979. "College Quality and Earnings," *Research in Human Capital and Development* 1: 103–121.
- Raudenbush, S. W., and A. S. Bryk. 1986. "A Hierarchical Model for Studying School Effects," *Sociology of Education* 59 (1986): 1–17.
- Reed, Ritchie, and Herman Miller. 1970. "Some Determinants of the Variation in Earnings for College Men," *Journal of Human Resources* 2: 177–190.
- Solmon, Lewis, and Paul Wachtel. 1975. "The Effects on Income of Type of College Attended," *Sociology of Education* 48: 75–90.
- Summers, Anita, and Barbara Wolfe. 1977. "Do Schools Make a Difference?" *American Economic Review* 67: 639–652.
- Wachtel, Paul. 1976. "The Effect on Earnings of School and College Investment Expenditures," *Review of Economics and Statistics* 58: 326–331.
- Wales, Terence. 1973. "The Effect of College Quality on Earnings: Results from the NBER-Thorndike Data," *Journal of Human Resources* 8: 306–317.
- Wise, David. 1975. "Academic Achievement and Job Performance," *American Economic Review* 65: 350–366.

The Financial Squeeze on Higher Education Institutions and Students

The Balance Between Quality and Access

W. Lee Hansen*
University of Wisconsin

This paper attempts to illuminate recent discussions about the tremendous financial pressures experienced by students, their parents, and colleges and universities in paying the costs of higher education (McPherson and Shapiro 1991). It does this by placing these developments in the context of long-run pendulum-like swings in society's interest in promoting greater access to higher education and enhancing the quality of the higher education enterprise. These swings are made apparent by using a new approach to organize and analyze the data on higher education finance.¹

The conclusion that emerges from this analysis is that we are currently in a transitional phase, following a thirty-year period of conflict between proponents of access-equity and of instructional quality. This shift in emphasis toward a joining of quality and access concerns is accompanied by an intense struggle over how the costs of achieving these objectives are to be shared among students, their parents, state and local taxpayers, voluntary contributors, and in the case of student financial aid, higher education institutions and the federal government (Hauptman 1990a, 1990b).

We start by assuming that the goals of higher education are influenced by a wide variety of internal and external forces. Whatever these aims may be, they do not emerge exclusively or even principally from internal analysis, deliberation, and pressures. Rather they grow out of external forces and events. This pattern is reflected in the common practice among educators of moving toward new goals and pushing for increased levels of funding in the wake of external events, such as renewed pressure for increased institutional support after Sputnik, or

---

* This paper is part of a larger collaborative effort with my colleague Jacob O. Stampen.

new student financial aid programs after the beginning of the War on Poverty. Each episode is followed by some new event that sets off a reaction in yet another direction, so that the process repeats itself.

Precedent for this view emerges from the research of historians as well as scholars from other disciplines who have tried to capture alternating patterns of change in economics, history, politics, and the like. These analyses use terms such as "tensions," "cycles," "pendulums," "spirals," and "dialectics" to describe the patterns that are uncovered.² Observers generally agree about the nature and identity of these cycles, whose life spans average between twelve and seventeen years (Schlesinger 1986, p. 24). They also agree that these cycles alternate between emphasizing public action versus private interest. These oscillations have been described by Hirschman (1982) as the "frustrations of public life" and by others as "liberal versus conservative" eras. Whatever the term, the meaning is generally the same.

The most active exponent of the cycles view is Arthur Schlesinger, Jr., who notes that each cycle "must flow out of the conditions and contradictions of the phase before and then itself prepare the way for the next recurrence."³ Schlesinger's analysis provides a useful framework for sharpening our research questions concerning recent changes in higher education goals and financing. The principal questions guiding this analysis are: First, how have the goals of higher education changed over the past half century? Second, does investment in higher education respond to changes in these goals? Third, how did changes in the goals and investment in higher education affect quality and access, the sharing of the costs of quality and access between students and society, and the ability of students and their families to finance college attendance?

Two sources of information are at hand to help answer these questions. One is the abundant literature on higher education. That literature can be distilled to reveal broad trends and critical shifts that illuminate the goals and direction of higher education. The results that emerge from such an analysis are difficult to assess because of the varying interpretations that can be given to them. The other is national statistical data on higher education enrollments, expenditure patterns, and the like. Such data reflect both the trends and responses to them just mentioned. The statistical data available for identifying finance-related changes are not ideal. Routinely gathered federal statistics on

higher education finance are incomplete, lack adequate detail, and suffer from definitional changes over time. These problems make it difficult to document in consistent fashion the financial trends as well as systematic changes in these trends.

However, even sometimes difficult to interpret information and imperfect data can yield important insights when the patterns of change can be related to the forces that underlie them. Only by trying to establish such connections is it possible to say something useful about the current policy debate on quality and access in higher education.

Cross Currents in Higher Education

American higher education over the past half century has been buffeted by a combination of demographic, social, political, and economic forces. Some of these forces are separable whereas others are closely linked. The linking of these forces may have been most obvious in the 1930s and 1940s. During the depression of the 1930s, college enrollments grew more slowly than they had in the past, and with the beginning of World War II they dropped substantially. Immediately after the war enrollments shot upward as a direct response to the GI Bill. Another view is that much of this gain served to "make up" for the slower enrollment growth of the 1930s and early 1940s.

After World War II, demographic factors emerged as a stronger element for change. By the early 1950s, most of the World War II veterans had passed through the educational system. Enrollment levels remained relatively stable until the late 1950s, due to the slow growth in the size of the traditional college-age population, and gradually increased into the early 1960s. An explosion of enrollments occurred in the mid-1960s as the post-World War II "baby boom" population reached maturity. Enrollments rose even more sharply, as interest heightened about increasing the enrollment of previously under-represented ethnic minorities as well as women. This increase continued through the 1970s, although the rate of growth slowed considerably. By the early to mid-1980s overall enrollment growth came to a virtual halt, and remained relatively unchanged for a few years; recently it has

renewed its upward climb. Enrollment declines attributable to the declining size of the traditional college-age population were offset by increased college attendance among people age twenty-five and above. Meanwhile, college participation rates for most minority groups have declined since the mid-1970s, as they have for males generally; at the same time, significant gains occurred for females.

Political forces have also exerted a powerful influence on the growth of higher education in the United States and are revealed most immediately in governmental actions. Ultimately, however, these actions reflect even more powerful forces, namely, the changing priorities of the citizenry who determine the focus of political action and availability of resources for higher education. The need to compete with the Soviets after Sputnik helped expand state and local resources for higher education. The same was true of concerns about broadening access for minorities and economically disadvantaged in the late 1960s and early 1970s. These efforts proved effective in galvanizing public opinion and bringing about the allocation of more federal resources to higher education. The student unrest experienced by higher education in the late 1960s and early 1970s probably had the opposite effect. Whether the current view that higher education can be an effective instrument for enhancing our international competitive situation—which would thereby increase the resources allocated to higher education—is valid or not remains unclear.

The economic environment also plays a key role in the shaping of higher education. Periodic wars and recessions have affected the tax revenues of the federal as well as state and local governments and have also had an impact on private contributions. As a consequence, the resources available to higher education institutions have fluctuated in often unpredictable ways. More important, competition from other state and local programs has reduced the relative allocation of resources to higher education. The productivity slowdown that began in the early 1970s made conditions even worse.

Though external forces are critical, it is also apparent that higher education has sought to chart its own course. Such efforts are reflected in a long series of reports that articulate the goals and aspirations of academic institutions.⁴ Closely related are the periodic attempts made by economists, historians, and other social scientists to offer new ideas and interpretations that stir the air and stimulate thinking about the

course of higher education—among them the current debate about diversity and cultural values.

Periods of Analysis

To facilitate this analysis we have defined four distinct periods which emerge out of our review of the qualitative material. The first period begins in the late 1930s and continues into the early 1950s. It reflects growing concerns about access to college, culminating with the GI Bill and its enormous impact on enrollment after World War II.

The second period begins in the mid-1950s and continues to the mid-1960s, thus capturing the enormous expansion of the higher education sector. It also picks up the emphasis on the elusive dimension of quality that was spurred by concern about America's lagging technology in the face of the Soviet launching of Sputnik. In addition, it reflects the widely publicized studies by economists establishing the link between investment in education and economic growth.

The third period, from the mid-1960s to 1980–81, embraces the search for ways to expand opportunities for students to attend college. The first phase began with the initiation of federal student aid programs in 1965 and culminated with the federal decision in 1972 to establish a national need-based student aid system. It was followed in the late 1970s by what can best be described as a phase of consolidating the financial aid system and confronting other equity-related problems, as exemplified by the Middle Income Student Assistance Act of 1978.

The fourth period began in 1980–81 and continues to the present. It represents the beginning of a sharp swing away from access to concerns about the quality of instruction, efficient use of resources, and once again education's role in economic growth. At present we may be entering a new phase, as concerns about access compete more actively with the push to improve quality.

These periods and their alternating swings between quality and access closely correspond to Schlesinger's pendulum-like political cycles mentioned earlier. Since the late 1940s, when society promoted increased college attendance, higher education sought to expand

access. When society promoted economic development and national security, higher education sought to improve quality.

The resulting swings do not necessarily emerge as sharply as the political cycles approach would suggest. Thus, they cannot be precisely dated in every case. Moreover, the data reflect the aggregation of not only changes in societal attitudes and behavior but also the perceptions of change emerging within higher education institutions and the actions these perceptions generate. Here we must ignore these micro-level underpinnings of these changes, even though they constitute an important part of the story.

Analytical Framework

With the time periods for this analysis established, we turn to the data in hopes of learning whether changing political-social-economic conditions and the accompanying societal mandates exerted any effect on resource allocation in higher education. We begin by describing the structure of the nation's investment in higher education institutions and in student support. We then examine higher education expenditures and revenues in an effort to highlight major trends and the interplay between the external and internal forces affecting resource allocation within the higher education sector. This information paves the way for measuring the burden of higher education costs and how these costs are shared among students/parents, state and local taxpayers/private donors, and also federal taxpayers, through federal student financial aid programs.

Our first task is to define proxy measures for the concepts of quality and access in the context of higher education finance. For purposes of this analysis, instruction-related costs are viewed as an indicator of efforts to promote quality. Tuition and fees, less student financial aid funds, are viewed as an indicator of efforts to improve access.

We recognize that these magnitudes are at best crude proxies for what we really want to measure. Rather than total student aid, we would prefer to focus on the portion of aid that enables young people from lower income families to undertake and continue with their higher education; in the absence of such aid, they would not be able to

do so. Similarly, rather than concentrating our analysis on all instruction-related expenditures, we would prefer to focus on the portion of those expenditures that “makes a difference” in quality (i.e., that produce greater and more lasting increments of student learning).

Even more important is the extent to which changes in these categories of expenditure affect quality and access. Spending more or less would obviously change the dollar totals. Whether, for example, additional expenditures would enhance quality or improve access is more difficult to say, given the complexity of higher education management.⁵ Nonetheless, for purposes of this analysis we shall take the dollar totals and changes in them as crude indicators of the relative priority given to quality and access in higher education.

If we are to determine who pays the instruction-related costs of higher education, then it becomes essential to identify how these costs are split between students and others. Thus, we must separate that portion of the costs paid by students through tuition and fees from that paid by taxpayers and private donors. The portion of instruction-related costs not paid by students is described as the nonstudent share, i.e., total expenditures paid by taxpayers for public institutions, and by voluntary contributions for private institutions. It should be obvious that there is no fixed distribution of these costs; their sharing can easily shift as conditions change.

The sharing of costs has still another dimension. It concerns the extent to which the share of instruction-related costs paid by students is offset by student financial aid. If we think of tuition and fees as the gross share of institutional costs paid by students, we can describe the net share as tuition and fees less student financial aid. The smaller the net share of total instruction-related costs paid by students, the greater the emphasis on access.

The Data

We rely heavily on official data from the Department of Education and its predecessor, the U.S. Office of Education. Because of changes in the data collection systems as well as periodic alterations in the definitions of expenditures and revenues, the detailed data are not com-

pletely comparable over the almost 50-year period under study. Nonetheless, the broad categories employed here are generally consistent. We caution readers that this analysis for all of higher education obscures potentially important differences between public and private-independent institutions; these differences will be examined in a subsequent paper.

We also utilize data on student financial aid. With the development of state-based student aid programs in the late 1950s, federal funding under the National Defense Education Act of 1958, and the major financial aid programs of the federal government beginning in the mid-1960s, these additional resources, which for the most part go directly to students, are not fully captured in the institutional data. To remedy this defect, we draw upon data on student aid expenditures compiled by the College Board beginning in the early 1970s. We have extended these data back to the late 1930s, and in the case of veterans' benefits provided through the GI Bill, back to the mid-1940s.

An unresolved problem with the financial aid data lies in figuring out how to eliminate from the totals those funds distributed to students attending proprietary schools. Such schools, and there are many more of them than there are colleges, are not included in the institutional data on expenditures and revenues. For this reason, the student financial aid data overstate the resources devoted to broadening access. This overstatement may have grown to as much as 15 percent of the total since the 1970s, as eligibility for student aid was expanded beyond higher education to include all of postsecondary education. (Work is underway to separate out student aid expenditures for students attending proprietary institutions).

The total value of resources for higher education is best captured by institutional data on expenditures shown in table 1 and by the College Board data on the amounts of aid provided to students shown in table 2. One difficulty arises with these data; serious overlap exists between the "scholarships and fellowships" item in the institutional data and the "institutional and other grants" item in the College Board data. Because the data are not quite comparable, we proceed under the assumption that the amounts of scholarships and fellowships shown in the institutional data are correct, and that the College Board totals are accurate. This requires subtracting the total of scholarships and fellow-

ships from total institutional spending before attempting to aggregate these two sources of data.

Table 1. Alternative Measures of Expenditures by Institutions of Higher Education, 1988-89 (in billions)

Type of expenditure	Current fund	Educational and general	Instruction-related
Instruction	\$38.8	\$38.8	\$38.8
Academic support, including libraries	8.9	8.9	8.9
Student support	5.8	5.8	5.9
Institutional support	11.5	11.5	11.5
Operation & maintenance of plant	8.7	8.7	8.7
Mandatory transfers	1.5	1.5	1.5
Public service	4.2	4.2	---
Research	11.4	11.4	---
Scholarships & fellowships	5.9	5.9	---
Auxiliary enterprises	12.3	---	---
Hospitals	11.8	---	---
Independent operations	3.0	---	---
Total	\$123.9	\$96.8	\$75.3

SOURCE: U.S. Department of Education, unpublished data.

Before continuing, it is helpful to know the overall level of resources devoted to quality and access. In 1988-89 (the most recent years for which complete institutional data are available) total current fund expenditures reached \$123.9 billion. Total expenditures on student financial aid reached \$25.5 billion. The total resources devoted to quality and access add up not to the sum of these two numbers, which

is \$149.4 billion, but rather to \$143.5 billion; this makes allowance for the \$5.9 billion "overlap" mentioned above (see table 3). Overall, these expenditures represent 2.93 percent of gross domestic product (GDP).

This figure is misleadingly high because total current fund expenditures include an array of activities that are not directly related to the instructional activities of colleges and universities. A closer approximation to the costs of interest for this analysis is provided by what are called educational and general expenditures. This amount is arrived at by subtracting from total current fund expenditures the costs of operating auxiliary enterprises, hospitals, and independent operations, all of which are activities bearing little or no direct relationship to the instructional missions of colleges and universities. The result is that educational and general expenditures, which in 1989-90 totaled \$96.8 billion, are 22 percent lower than the total current fund expenditures (see table 3).

Table 2. Financial Aid Expenditures for Postsecondary Education, 1988-89 (in billions)

Type of student aid	Amount
Federal supported programs	
Generally available aid	
Grants, loans, work study, institutional aid	\$18.4
Specifically directed aid	
Veterans, military, etc.	1.5
State grant programs	1.6
Institutional and other grants	4.0
Total	\$25.5

SOURCE: The College Board, *Trends in Student Aid: 1981 to 1991*, August 1991, Table 1.

NOTE: The amounts shown above include some aid awarded to students attending proprietary schools which are not included in the data for institutions of higher education. Hence, the student aid data overstate the amounts of aid available to college students. The magnitude of the overstatement is in the 15 percent range.

While educational and general expenditures come closer to the mark, they still include activities that go well beyond instruction. Two types of expenditures need to be excluded. One is for public service activities directed to external audiences; included would be such things as extension activities carried on by land-grant institutions. The other

is for research, a central activity of major research universities and typically carried out with the help of external funding. The fact that research produces new knowledge, some of which filters back into instruction through the teaching done by researchers and through the professional journals and textbooks used by countless students across all types of colleges and universities, suggests that some part of research activity is instruction-related. Because it would be so difficult to assess the impact of research on instruction for undergraduates in particular, no attempt is made to allocate any part of research expenditures to instruction.

Table 3. Overall Institutional and Student Financial Aid Expenditures on Higher Education, 1988-89 (in billions)

Expenditures	Institutional data		Student financial aid expenditures (3)	Total expenditures unduplicated (4)	Total expenditures as percent of GDP (5)
	Total (1)	w/o SFA (2)			
Current Fund	\$123.9	\$118.0	\$25.5	\$143.5	2.93%
Education & general	96.8	90.9	25.5	116.4	2.38%
Instruction-related	N/A	75.3	25.5	100.8	2.06%

SOURCE: Calculated from published and unpublished data from the U.S. Department of Education and the College Board.

NOTE: Column 4 is sum of columns 2 and 3.

After excluding expenditures on public service and research (see table 1), we arrive at instruction-related expenditures (shown in line 3 of table 3) which in 1988-89 amounted to \$75.3 billion. When combined with the student aid total, we find that expenditures of \$100.8 billion on quality and access represent 2.06 percent of GDP (table 3). To provide a point of comparison, total current expenditures on K-12 education accounted for 4.2 percent of GDP.

We also need to know the amount of tuition and fees paid by students. This information comes from the institutional revenue data. In 1988-89 tuition and fee revenue amounted to \$30.8 billion. To the extent that instruction-related expenditures amounted to \$75.3 billion, the tuition and fees component of revenue covered 40.9 percent of these costs. The remaining revenue used to pay instruction-related costs is provided largely by state and local governments in the case of

public colleges and universities and by private donors in the case of private institutions. While both instructional costs and tuition rates differ appreciably among public and private institutions, these differences are ignored here.

Normalizing the Data

Before moving ahead with the analysis, the data must be normalized in order to facilitate comparisons over time. Instruction-related expenditures must be adjusted for changing enrollment levels. This is done by constructing a new measure, instruction-related expenditures per full-time equivalent (FTE) student. It is important to use FTE enrollment because of the sharp increase in the proportion of part-time students since the 1970s.

To assess the strength of efforts to enhance both quality and access we need some standard against which to make comparisons. The ideal would be a measure of changes in the relative capacity of the economy to finance quality and access in higher education. Such a measure makes it possible to avoid having to correct for price level changes because it converts the data from nominal to relative values.

Since GDP provides such a convenient and well-understood measure of aggregate output and hence aggregate capacity to pay, we utilize GDP per employed member of the civilian labor force (CLF) as an indicator of the public's capacity to pay. GDP is preferable to other widely used measures because it reflects the value of all goods and services produced in the economy; it can also be related more directly to frequently made comparisons of higher education expenditures. Thus, GDP per member of the CLF provides a rough measure of the ability of the average worker to provide tax and nontax support for higher education.

The final step requires us to express the various cost measures, such as instruction-related cost per FTE student, as a percent of GDP per member of the CLF. With these measures it becomes possible to highlight relationships among the level of instruction-related costs, who pays for them, and how financial aid affects the student share of these

costs. This information makes it possible to offer a preliminary assessment of society's efforts to promote quality and access simultaneously

The Results

The key measures needed for this analysis are expressed as a percent of GDP per member of the CLF and are presented in table 4. The first column shows instruction-related costs per FTE as a percent of GDP/CLF. These costs rose steadily from 1947–48 to 1972–73, dropped in 1976–77, began increasing again after that, and by 1988–89 exceeded the previous high in 1972–73. Within the framework presented here, it appears that investment in quality increased steadily through the early 1970s, dropped off a bit later in the decade, and then began rising again. The rise in the 1980s proved to be steep, when emphasis once again shifted to improving the quality of higher education.

The access story is more difficult to follow because of its several distinct components. The first is the pattern of change in tuition and fees. The second is institutional aid, which colleges and universities provide out of their own resources. The third is other student aid, which comes largely from veterans' benefits, social security benefits for eligible college students, and federal student aid programs.

Further clarification is necessary concerning these three sources of other student aid. Veterans' benefits provided a major stimulus to college attendance immediately after World War II, again but to a lesser extent after the Korean War, and yet again but to an even smaller degree after the Vietnam War. The benefits available to World War II veterans included a monthly stipend plus government payment of all tuition and fees. The fact that the "other aid" was so great right after the end of World War II is not surprising; approximately half of all college students at the time were veterans. Their benefits included government-paid tuition of up to \$500 per year and a monthly allowance which for a single veteran without dependents provided \$65 per month. The impact of veterans' benefits diminished through the 1950s because fewer Korean War veterans attended college under a somewhat different GI Bill set up to deal with this new group of veterans.

Under this legislation veterans were not reimbursed for their tuition and fees, though the monthly stipend for a single veteran had risen to \$105 per month. By the late 1950s the amount of funding provided through such benefits had greatly diminished. This aid increased again in the late 1960s and early 1970s as Vietnam War veterans enrolled and was based on GI Bill benefits similar to those given to Korean War veterans.

Table 4. Instruction-Related Costs, Tuition and Fees and Student Financial Aid Per Full Time Equivalent (FTE) Student Relative to Gross Domestic Product (GDP) Per Member of the Civilian Labor Force (CLF) (in percent)

Academic year	Instruction-related costs (1)	Tuition and fees (2)	Institutional aid (3)	Other aid (4)	Total student aid (5)
1947-48	13.2	3.5	0.5	12.5	13.0
1951-52	14.5	4.3	0.7	6.4	7.1
1957-58	15.3	5.3	0.7	2.4	3.1
1965-66	16.8	6.3	1.0	0.9	1.8
1972-73	18.4	6.2	1.4	4.9	6.2
1976-77	17.8	6.1	1.2	5.8	7.0
1980-81	18.1	6.3	1.1	6.5	7.6
1984-85	19.1	7.3	1.3	5.2	6.5
1988-89	20.1	8.2	1.6	5.2	6.8

SOURCES: Calculated from published and unpublished data from the U.S. Department of Education and the College Board. Data on GDP and CLF are from various issues of the Economic Report of the President.

NOTE: Calculations for years prior to 1965-66 are based on GNP rather than GDP.

Since the shift to an all-volunteer army in the early 1970s, it has been more difficult to view veterans' educational benefits as a form of student financial aid. Instead, such benefits can be considered a part of the military compensation package, a sort of deferred wage payment granted in the form of educational benefits. Another argument for not including veterans' benefits in student aid is that these benefits to veterans of World War II, Korea, and Vietnam represented an effort by society to make up for the well-below market wages paid to the many

men who had been drafted into military service. For the purpose of this analysis, however, veterans' educational benefits are viewed, as they are by the College Board, as a component of student financial aid.

Social security benefits for eligible dependents began in 1965 and were finally phased out in the early 1980s. These benefits are more problematic because they were confined to college students age eighteen to twenty-one. After the establishment of need-based Pell Grants in 1972, the rationale for continuing social security benefits was seriously undermined. It took a decade before Congress finally voted to eliminate them.

More important than the aid provided by institutions, at least since the mid-1960s, is that offered by the federal government through grants, loans, and work-study programs. The development of student aid programs dates from 1964 when anti-poverty legislation established work-study programs, and a year later when the Higher Education Act of 1965 established the Guaranteed Student Loan program and a series of related institution-based aid programs. This was followed by another major initiative in 1972, when Congress passed legislation to create what are now called Pell Grants.

The data on institutional student aid, other student aid, and total student aid appear in columns (3), (4), and (5) of table 4. The results are expressed as aid per FTE student as a percent of GDP/CLF. Institutional aid grew sharply through the early 1970s. Thereafter, the percentage remained roughly constant through the middle 1980s, when it increased quite sharply. Other aid varied more widely in response to changes in the level and mix of veteran's benefits and federal student aid programs. The precipitous drop from 12.5 percent in 1947-48 to 0.9 percent in 1965-66 is a result of the drying up of veterans' benefits. So also is the sharp increase by 1972-73 as federal student aid programs expanded and veterans' benefits expanded once again. Federal aid continued increasing to 1980-81. Since then other aid declined, falling back close to its 1972-73 level.

The pattern of change in total aid is dominated by movements in other aid. Nonetheless, changes in institutional and other aid may move together or in opposite directions. Since 1980-81 the decline in other aid was partially offset by increased institutional aid. Some would argue that the decline in other aid pushed institutions to provide more aid from their own budgets. Another explanation is that increased

student aid from institutions represented an effort to ameliorate the sharp increases in tuition and fees that were then taking place.

The impact of student aid on access is revealed in table 5. Column (2) shows tuition and fees that can be described as the gross student share of instruction-related costs. Column (3) shows the net student share, which is tuition and fees less institutionally-provided student financial aid. Column (4) shows what can be called the net net student share, which is tuition and fees after subtracting both institutional and other student aid. Negative values in column (5) indicate that total student aid exceeded total tuition and fees paid by students, whereas positive values indicate the opposite.

Table 5. The Burden of the Costs of Higher Education; Based on Costs Per Full-Time Equivalent (FTE) Student Relative to Gross Domestic Product (GDP) Per Member of the Civilian Labor Force (CLF) (in percent)

Academic year	Instruction-related costs (1)	Gross student share: tuition and fees (2)	Net student share, incl. inst. aid (3)	Net net student share, incl. all aid (4)
1947-48	13.2	3.5	3.0	-9.4
1951-52	14.5	4.3	3.6	-2.8
1957-58	15.3	5.3	4.6	2.2
1965-66	16.8	6.3	5.3	4.5
1972-73	18.4	6.2	4.8	-0.0
1976-77	17.8	6.1	4.9	-0.9
1980-81	18.1	6.3	5.1	-1.3
1984-85	19.1	7.3	6.0	0.8
1988-89	20.1	8.2	6.6	1.4

SOURCES: Calculated from published and unpublished data from the U.S. Department of Education and the College Board. Data on GDP and CLF are from various issues of the Economic Report of the President.

With this as background, we come back to the quality-access trade-off. With respect to quality, the increasing figures from 1947-48 through 1972-73 (column (1) of table 5) suggest that quality was ris-

ing. This rise was followed by a decline that continued through the early 1980s. However, the trend has been upward since the late 1970s.

Meanwhile, the focus on access was exceptionally strong in the late 1940s and early 1950s. It dropped off sharply through the late 1950s and continued doing so into the middle 1960s. With passage of the Higher Education Act of 1965, the pattern suddenly reversed itself, as evidenced by a sharp fall in the net net-student share, to zero in 1972–73 and even lower through the remainder of the decade. Since 1980–81 the aggregate amounts invested in student aid have fallen short of total tuition revenue.

The resulting pattern can be summarized as follows:

Periods	Quality	Access
WW II to 1947–48	Presumably high	Rising
1947–48 through 1965–66		Falling
1947–48 through 1972–73	Rising	
1965–66 through 1980–81		Rising
1972–73 to 1980–81	Falling	
1980–81 to 1988–89		Falling
1976–77 to 1988–89	Rising	

In general, when the emphasis on access falls, the emphasis on quality rises, and vice versa.

Sharing the Costs

How are the costs of achieving quality and access being shared? Table 6 can help answer this question. One view of this sharing is provided by columns (1) and (2), which indicate the division of instruction-related costs between students and others—meaning mostly taxpayers for public institutions and voluntary contributors for private institutions. The student share rose steadily through 1965–66, dropped off a bit and then remained relatively constant through 1980–81, and afterward increased once again to its highest level ever. The magnitude

of the increase rose from about one quarter to slightly more than 40 percent of instruction-related costs.

Table 6. Sharing the Costs of Higher Education, Based on Costs Per Full Time Equivalent (FTE) Student Relative to Gross Domestic Product (GDP) Per Member of the Civilian Labor Force (CLF) (in percent)

Academic year	Gross student share tuition and fees (1)	Non-student share: taxpayers institutional and donors (2)	Other aid (3)	Total aid (4)	Aid (5)
1947-48	26.3	73.7	3.5	94.6	98.1
1951-52	29.7	70.3	4.7	43.9	48.7
1957-58	34.6	65.4	4.9	15.6	20.4
1965-66	37.5	62.4	5.8	5.2	10.9
1972-73	33.6	66.4	7.3	26.4	33.7
1976-77	34.2	65.8	6.7	32.5	39.2
1980-81	34.6	65.4	6.3	35.8	42.1
1984-85	38.0	62.0	6.6	27.3	33.9
1988-89	41.0	59.0	7.9	26.1	33.9

SOURCES: Calculated from published and unpublished data from the U.S. Department of Education and the College Board. Data on GDP and CLF are from various issues of the Economic Report of the President.

The reasons why the student share increased so dramatically need to be examined. Several explanations come to mind. One is that it may have been politically more difficult to increase nonstudent assistance than student contributions. When revenue is tight because of rising demands for other publicly provided goods and services and the reluctance of taxpayers and donors to provide more funds, it is easier to increase the tuition of already-enrolled students who, because of the large economic benefits of college looming ahead, sense that they must pay. Another plausible explanation is that because private benefits to college attendance are so apparent while the social benefits are more difficult to document, society has been moving to require students, the most direct beneficiaries of college, to pay an ever larger share of the

instructional costs. These and other possible rationales obviously require more careful study.

Another way to examine the sharing of these costs is to compare instruction-related expenditures with institutional and other aid, as seen in table 6. Column (3) shows that institutional aid has always provided a relatively small share of total instruction-related expenditures. Interestingly, institutional aid increased steadily through 1972–73 and then dropped off, no doubt because of the growth of federal student aid. However, institutional aid resumed its steady increase from 1980–81 through 1988–89, as institutions allocated to student aid more of their additional revenue from tuition and fees.

The patterns of change in other aid and total aid are similar to those shown in tables 4 and 5. Total aid about equalled total instructional costs in 1947–48 but then fell to almost nothing by 1965–66. With the beginning of federal student aid programs in 1965–66 a sharp increase occurred, which continued through the 1980s. Since then other aid dropped, largely as a result of the slow growth of federal student aid funds.

Interpretation/Summary and Discussion

In examining the goals and financing of higher education over the past half century, we find cyclical patterns of change. These changes reflect cycles similar to those noted by Schlesinger, cycles that may also exist in other areas of economic activity. For higher education, however, these cycles translate essentially into two alternating mandates, one to improve quality and the other to improve access. Such cycles can be viewed as representing normal variation within the system.

Over the period since World War II, the rate of investment in higher education has risen considerably. As shown in table 7, investment rose from less than 1 percent prior to the middle 1960s, when it first exceeded 1 percent; since 1972–73 it has been stabilized at 1.5 percent. Much of the increase came from the expansion of higher education enrollments which more than quadrupled. As a percentage of the civilian labor force, the number of FTE students slightly more than dou-

bled. Overall, quality increased as shown earlier in table 4, with instructional costs rising from 13.2 to 20.1 percent. This is an impressive gain, occurring as it did when enrollment increased so dramatically. Thus, quality and access improved substantially over the period as a whole.

Table 7. Indicators of Expansion of Investment in Higher Education

Academic year	Instruction- related expenditures as a percent of GDP (1)	All student aid as a percent of GDP (2)	FTE enrollment (in millions) (3)	Total investment in higher education as a percent of GDP (4)	FTE enrollment as a percent of CLF (5)
1947-48	0.49	0.48	0.07	2.3	3.7
1951-52	0.45	0.21	0.66	1.0	3.1
1957-58	0.60	0.12	0.72	2.6	3.9
1965-66	1.04	0.11	1.15	4.7	6.2
1972-73	1.49	0.50	1.99	7.3	8.1
1976-77	1.50	0.59	2.09	8.1	8.4
1980-81	1.47	0.62	2.09	8.8	8.1
1984-85	1.48	0.50	1.98	9.0	7.8
1988-89	1.54	0.53	2.07	9.5	7.7

SOURCES: Based on data from U.S. Department of Education and Economic Report of the President.

What we find particularly interesting is how changes in the goals of higher education affected quality, access, and the sharing of costs between students and society. The relative constancy until recently in the gross student share, represented by tuition and fees, and the systematic changes in instruction-related costs and the net student share, are remarkable. The fact that these latter two measures displayed such variation is an interesting commentary on the changing priorities in higher education finance. Equally surprising is the fact that total student financial aid exceeded combined tuition and fee revenues in two quite different time periods—through most of the 1970s and also much earlier, just after World War II.

If the late 1960s and 1970s was a period of concern about access, the concern of the 1980s was with quality. By the measures adopted for this analysis, quality declined in the 1970s and increased in the 1980s, whereas access increased in the 1970s and decreased in the 1980s. It should be noted, however, that increased investment in quality in the 1980s was small relative to the increase in access from the mid-1960s to the mid-to late-1970s. As a result, little has materialized in the way of quality gains.

Throughout the 1970s the push for wider access through increased student aid brought with it pressures to hold down tuition increases. As a result, additional demand for higher education was stimulated, which brought enrollments to even higher levels in a period when constrained budgets made it increasingly difficult to hire additional faculty. As support for instruction-related costs lagged, the principal casualty was faculty salaries, which fell dramatically in real terms through the 1970s and into the early 1980s (Hansen 1986).

By the early 1980s the results of this process were becoming more evident. Though increased student aid may have helped stimulate enrollments, it was not clear that it had done much to stimulate the enrollment of young people from lower income families.⁶ Nonetheless, institutions needed more resources to hire faculty in an ever tighter labor market. As faculty salaries rose, instructional costs began to climb. Simultaneously, student aid resources contracted in relative terms.

The 1980s saw the absence of increases in traditional forms of financial support, which meant that tuition and fees had to be raised. To deal with the hardship created by this response, institutions began providing additional financial aid out of their own resources. Increasingly, however, the resolve to continue this practice appears to be weakening. Despite the growing emphasis on quality, society's investment in it increased only slightly in the 1980s because overall resources for higher education remained tight.

During the 1980s, a shift in public and institutional priorities away from access and toward quality appeared to be underway. This move was financed largely by students through tuition increases rather than by traditional sources of support, such as state and local taxpayers and private donors. In an attempt to respond to the growing concern about quality, institutions have been forced to find whatever financial support

they could. In the absence of other support, tuition and fees were raised.

An unresolved question is how much the emphasis on access in the 1970s contributed to the nation's goal of enhancing equal educational opportunity. Indeed the net cost of college attendance declined sharply for young people with incomes low enough to qualify for student aid; this proved to be a major accomplishment. While college participation rates for low-income students did not increase, evidence for the early 1980s shows that low family income was not by itself an important determinant of whether students dropped out or completed college. The growing availability of financial aid largely offset the effects of low family incomes. Rather, weak academic preparation, as indicated by mediocre performance in high school and low scores on standardized tests, constitutes the most important remaining barrier to expanding access to college.⁷ This suggests that access will be difficult to increase without improving the quality of instruction at the secondary level. In other words, current efforts to improve the quality of instruction could be effective if in the process academic performance improves among high school graduates from low-income families. As larger proportions of better-prepared young people enter college, student financial aid may become even more effective as a means to ensure greater equality of opportunity in higher education.

Still another question concerns the impact of current efforts to improve the quality of education. The implicit argument is that tuition increases have been required to improve the quality of the education. By paying higher faculty salaries, increasing expenditures to update equipment and facilities, and introducing new technology to the classroom, institutions believe they have been improving quality. Most institutions would have preferred to find other ways of meeting these increased costs; they would have liked to receive more state and local revenue as well as larger voluntary contributions. Despite the much-publicized fact that tuition and fees have increased sharply, public reaction against these increases has not been noticeably strong. It has certainly not been strong enough to elicit additional support from other sources or to restore the real levels of faculty salaries. Whether these changes have adversely affected quality remain to be determined.

The challenge now lies in finding better ways of using existing resources, so as to continue to achieve increasing access and improving

quality. If this can be done, the chances for obtaining additional resources to broaden access and enrich quality should be greatly enhanced.

NOTES

1. Most analyses of the higher education finance data show relatively little in the way of systematic patterns of change.

2. The importance of cycles has been emphasized primarily by the Schlesingers: see Arthur M. Schlesinger, Jr. (1986) and Arthur Schlesinger, Sr. (1949). Also see McClosky and Zaller (1984), Kaestle (1972), Hirschman (1982), and Hegel (1817).

3. Schlesinger goes on to say that such cycles "cannot be determined, short of catastrophe by external events. War, depressions, inflations may heighten and complicate moods, but the cycle itself rolls on, self contained and self sufficient" (pp. 27-29). Hegel might have characterized a cycle as a part of a dialectical process wherein each asserts a thesis which, as time passes, draws opposition resulting in the formation of an antithesis, which causes the beginning of a new cycle. However, surviving elements of a previous cycle's thesis become permanent parts of a presumably richer and more highly developed array of public policies.

4. Those reports include the President's Commission on Higher Education, *Higher Education for American Democracy* (Washington, DC, Government Printing Office, 1947); The Report of the President's Commission on National Goals, *Goals for Americans* (Englewood Cliffs, NJ: Prentice Hall, 1960); Carnegie Commission on Higher Education, *Quality and Equity: New Levels of Federal Responsibility for Higher Education* (New York: McGraw Hill, 1968); National Commission on Excellence in Education, *A Nation at Risk: The Imperative for Educational Reform* (Washington, DC, Government Printing Office, 1983); Association of American Colleges, *Integrity in the College Curriculum: A Report to the Academic Community* (Washington, DC, Association of American Colleges, 1984); and, Ernest L. Boyer, *College: The Undergraduate Experience in America* (Princeton, NJ: Carnegie Foundation for the Advancement of Teaching, 1987).

5. The evidence indicates that over the past decade or more employment in higher education has increased at a much faster pace for nonfaculty than faculty personnel. Whether this represents an enhancement of instruction quality is doubtful. For more details, see Bergmann (1991).

6. For two different views, see McPherson and Shapiro (1991) and Hansen (1983).

7. These patterns are documented by Stampen and Cabrera (1986) and also Cabrera, Stampen, and Hansen (1990).

References

- Bergmann, Barbara R. 1991. "Bloated Administration, Blighted Campuses," *Academe* (November/December).
- Cabrera, Albert F., Jacob O. Stampen, and W. Lee Hansen. 1990. "Exploring the Effects of Ability to Pay on Persistence in College," *Review of Higher Education* 13 (Spring): 303–36.
- Hansen, W. Lee. 1983. "Impact of Student Aid on Access." In *The Crisis in Higher Education: Proceedings of the Academy of Political Science*, Vol. 35, No. 2, pp. 84–96.
- _____. 1986. "Changes in Faculty Salaries." In *American Professors: A National Resource Imperiled*, Howard R. Bowen and Jack Schuster, eds. Oxford: Oxford University Press.
- Hauptman, Arthur M. 1990a. *The College Tuition Spiral: An Examination of Why Charges Are Increasing*. New York: Macmillan.
- _____. 1990b. *The Tuition Dilemma: Assessing New Ways to Pay for College*. Washington, DC: Brookings Institution.
- Hegel, George. 1817. *Encyclopedia of the Philosophical Sciences in Outline*, 3d. edition, 1971. Oxford: Clarendon Press.
- Hirschman, A. O. 1982. *Shifting Involvements: Private Interest and Public Action*. Princeton: Princeton University Press.
- Kaestle, Carl. 1972. "Social Reform and Urban Schools," *History of Education Quarterly* 2 (Summer): 211–228.
- _____. 1990b. *The Tuition Dilemma: Assessing New Ways to Pay for College*. Washington, DC: Brookings Institution.
- McClosky, Herbert, and John Zaller. 1984. *The American Ethos: Public Attitudes Toward Capitalism and Democracy*. Cambridge, MA: Harvard University Press.
- McPherson, Michael S., and Morton Owen Shapiro. 1991. *Keeping College Affordable: Government and Educational Opportunity*. Washington, DC: Brookings Institution.
- Schlesinger, Arthur M., Sr. 1949. *Paths to Present*. New York: Macmillan.
- Schlesinger, Arthur M., Jr. 1986. *The Cycles of American History*. Boston: Houghton Mifflin.
- Stampen, Jacob O., and Alberto F. Cabrera. 1986. "Exploring the Effects of Student Aid on Attrition," *Journal of Student Financial Aid* 16, 2: 25–40.

The Economics of Education in a World Of Change

Mary Jean Bowman
University of Chicago
Emeritus

The Meanings and Scope of Education, Economics, and Change in the Present Context

Central to any meaningful analysis of the economics of education is the notion of change, an idea that I wish to explore in these pages. To do so it is first necessary to bring some clarification to the meaning and scope of the terms “education” and “economics of education.” Only then is it possible to determine what sorts of changes—societal, political, economic, or others—are indeed relevant to this context. I then turn to the implications of change as it affects and is affected by education in industrialized societies. Although they merit their own study, for want of space I refer only incidentally to common and distinctive aspects of such change in less developed countries.

First, education is much more than schooling. It is all sorts of investments in learning. This must be obvious as soon as one looks across diverse societies and cultures around the world at any given time. Even illiterate societies have educational systems. It is equally obvious if we look over historic time in given societies, whether in the Eastern or the Western hemisphere. Lawrence Cremin is well known for his broad definition of education, which encompasses all investments of time in learning. This notion, however, leaves open the question of how far schools in fact educate, and whether education (in schools or elsewhere) is always a “good.”

Second, economics is more than what money measures. If usually we think of “education” in normative terms, what about “economics”? Which type of economics—positive or negative—is primarily relevant here, and to what extent can the positive and the normative be separated?

This leads immediately to the question: Does the "economics of education" encompass broader values, or is it concerned only with monetary costs and returns? To be sure, most of the benefit-cost analyses of investments in education have been limited in this manner. So limited too, have been treatments of education in aggregative analyses of "national economic growth." But "human resources" are much more, in both individual and societal perspective, than potentials for contributing to monetary returns.

Third, societal change has many facets. We all experience change over a life span, but this could be the situation even in an essentially changeless, traditional society. Today's "world of change" has certain unique features. Here societal traits, some of which may shift rapidly, bring myriad changes that are manifest over a single life span of an individual. Other changes become evident in their impacts over longer periods.

Three societally relevant sorts of intra-cohort changes may be conceptually distinguished. (1) *General economic cycles* bring cyclical changes in both investments in education and returns on such investments. (2) *Rapid, innovative* changes can have immediate impacts on the demands for services of skilled people. Such changes may be technological or organizational or a combination of the two. (3) Finally, there are *education-induced* societal changes within the adult life span of a single cohort, which arise in response to changes in the distribution of education among a population, whatever the shares of overlapping cohorts in such a change. Stated to include demographic changes in age and sex distributions, they might better be termed changes in the distribution and quantity of human capital.

In fact, changes of all three of these societal sorts can and often will arise contemporaneously. Sorting out these components in changing associations between education and earnings has been one of the important endeavors in empirical studies of the economics of education in recent years.

Where change is so slow as to be barely perceptible within the span of an individual life, the immediate consciousness of change over a lifetime will reflect only age cycles that seem to repeat themselves. However, when change is cumulative over extended periods, whether slow or rapid, its analysis in relation to the economics of education has often been characterized merely by comparisons between historic eras

or sharply contrasting contemporary societies—relationships that are usually simplified by disregarding societal intra-cohort shifts. Such simplification can be well justified except in cases where rapid intra-cohort shifts constitute a major feature of an historical era or a particular contemporaneous society. This exception, however, is extremely important.

Change and the Economics of Education in Industrialized Countries

Concentrating on change and education in industrialized countries, five main subjects call for attention. First is the proposition that disequilibria drive modernization, and that human capital plays a major part in that process. Second, uncertainty in the face of change has implications of uncertainty for education, and in particular for the roles of general education in a world of change. Third must come consideration of the ongoing debates concerning vocational, specialized, and general education with rising affluence (pervasive in connection with less developed as well as economically advanced nations). Fourth, what may we have to say about the roles and distribution of postschool training among members of a population in the face of dynamic change? Finally, are those who drop out of school early irrational? What about motives and incentives for educational decisions and concerning postschool behavior?

The argument that human capital and disequilibria constitute the mainspring of growth is the theme of a 1990 book by T. W. Schultz, entitled *Restoring Economic Equilibrium*. There he stresses three “common omissions” in modern growth theory. These are (1) *specialization* as a key to most modern increases in income; (2) *disequilibria* as increasing incomes are realized from advances in technology, from the proliferation of human capital, and from other sources; and (3) *entrepreneurs as agents* in restoring equilibria. The emphasis on specialization is not unique to Schultz. Indeed, this enlargement of Adam Smith's division of labor has characterized a number of papers, published and unpublished, by other economists over the past decade. Nor is a stress on entrepreneurship new; it was central to Schumpeter's the-

ory of economic development eighty years ago, although it has received little attention recently. What is distinctive in Schultz's recent work is the extent to which he focuses on *small* entrepreneurs and the nature of his treatment of "disequilibria." Criticizing economists in general, he writes, "It has become an art to conceal economic disequilibria that occur as a consequence of modern increases in income," whether such increases arise from technological change or from growth in human capital. In Schultz's (1975) view, disequilibria caused by modernization are seen as signals of income-increasing processes, which in turn give rise to new opportunities—hence his emphasis on the importance of "the ability to deal with disequilibria." That ability, designated elsewhere as "allocative" versus "worker" ability, has been shown to be associated with the completion of higher levels of schooling in both the United States and India.¹ A partial appreciation of the importance of this sort of entrepreneurial ability appears frequently today in nonacademic publications—for example, in *Forbes* magazine and the *Wall Street Journal*. But there is a paradox in all this. It would seem that specialization yields progress, but that a general education should provide the firmest base for dealing with and adapting to change. Specialization precipitates the disequilibria that give rise to economic progress; general education underlies abilities to remove those disequilibria through their creative resolution.

Second, a society in which the unexpected is perceived as the norm calls for "general education." Change breeds uncertainty, and a world of dynamic change is inevitably characterized by doubts and questions that affect the economic logic of choices in preparation for and in reactions to the unexpected. In addition, there are uncertainties for the individual in an advanced market economy that overlap with changes in the societal environment. This raises the question: How far do individual uncertainties inherent in a market system coincide in the nature of their effects with the overlapping uncertainties associated with cyclical or rapid technological or structural changes?

One essentially simple theme of this discussion is the importance of ensuring flexibility in adaptation to changes in technologies and in skill demands and supplies. This leads to a fundamental proposition regarding educational choices as viewed from both individual and societal perspectives. In brief, a critical function of education in the early years would seem to be "general" learning, in that it will provide a flexible

foundation for future learning, whether in formal institutions or elsewhere.

But what is a "general education"? Most fundamental and most general of all is undoubtedly the learning of attitudes and behavior that takes place (or fails to take place) in the home. This is why advanced industrial societies have exerted ever increasing pressure to provide formally for the education of children in the preschool years. It is a reason also for tendencies to ask ever more of the schools as socializing agencies, demanding that they go far beyond their earlier roles (some church or elite boarding schools aside). Moreover, work experience in itself may provide elements of general education for future job success in almost any vocation. All of these tendencies can have important economic effects, even if we define "economic" in the narrowest, monetary sense.

Beyond this, numeric and verbal literacy are undoubtedly among the most elemental foundations of general education throughout the industrialized world. But definitions of "functional verbal literacy" change, while numerical literacy tends to become less demanding on the one hand (with the omnipresent cash register) and more demanding on the other hand, in terms of mathematical literacy. Cutting across them all now is the issue of computer literacy.

Meanwhile, "practical"—not to be confused with "vocational"—learning has been coming in for more attention, along with estimates of its costs even if not yet of its returns. What constitutes "generally practical" learning will depend on environmental conditions.

It is evident that the more technologically advanced and diverse a society, and the more rapid the pace of economic change, the greater must be the demand for a general education that can foster adaptability, whether from an individual or a societal perspective.

Third, where, then, does specialization come in? A fallacy that remains common in some quarters, even today, is the notion that schools should "turn out" students fully trained for particular intermediate-level jobs. Usually this argument underlies demands for the vocationalization of secondary schools. But frequently it confounds the vocational with the practical, which may be of general relevance to most members of a population. Even if the "practical" skill is of general relevance, questions may still arise as to the cost and effectiveness of providing it in schools or through other channels. This leads into the

question, What level of general schooling should candidates attain before vocational specialization? Cogent arguments for increased specialization may well hold with reference to postcollege education, for example, even while in a particular environment arguments for widespread vocational specialization during secondary school might attract little support. What we have to remember is that the case for rising specialization in advanced economies rests on two assumptions: (1) that the students already will have attained high levels of general education, and (2) that they are well prepared to deal with changes in future knowledge and practice in the general area of their specialization. Rapid change in an advanced society supports and depends upon both high levels of training in general competence in a cluster of specialties *and* increased high-level specialization.

It is in such a context that Rosen (1983) contributed his eminently readable essay on specialization, the gist of which was that incentives to specialize arise from increasing returns in utilization of human capital. This comes about because of the indivisibility of human capital, embodied as it is in the human being. In Rosen's words, "The return to investment in a particular skill is increasing in its subsequent rate of utilization because investment costs are independent of how acquired skills are used" (p. 44). He illustrates the working of this principle by pointing to the differences between men and women in incentives to invest in human capital, and to the division of labor within households. Becker and Murphy (1990), among others, have expanded on Rosen's discussion to carry further the argument concerning the importance of rising specialization. As an economy becomes technologically more complex and the quantity of disembodied knowledge in a society becomes progressively larger, no one person can contain more than a minute fraction of the total. Limitations on the extent of specialization go beyond Adam Smith's size of the market to include the costs of coordination. In our day we are witnessing a multiplication and refinement of communication technologies that lowers coordination costs and the barriers of distance, even as an increasingly complex market economy takes over a major part of the task of coordinating the work of ever more specialists.

An argument between T. W. Schultz and his Chicago colleagues persists with respect to the concept of disequilibria and the place that Schultz has given to that concept. There is substantial agreement,

nonetheless, with respect to the importance of high-level specialization in the dynamics of an advanced economy.

Next is the question: How do the uncertainties of change and postschool human investments interact? In an economically advanced and dynamic society, continued learning over the adult years must be important for almost everyone but will not be the same for all. For some it may be in large part a process of recovery from earlier mal-allocation of time away from what should have been learned, even at relatively low levels of both general capabilities and particular vocational skills. At the other extreme, manifest in most professions, continuous intensive learning is required merely to keep up with rapid increases in knowledge. At both extremes, and in between, postschool learning is an essential ingredient in sustained productivity for both individuals and society. Catching up and keeping up both are important, whether or not entrepreneurial in an innovative sense. Even if catching up and keeping up are in themselves more reactive than creative, both are essential in the processes of societal change.

Indeed, whether a society is characterized by dynamic change or not, an examination of postschool learning is necessary in order to identify returns on investments in schooling, insofar as the extent of postschool investment in human capital is associated with the extent and nature of prior schooling. Or to be more precise, such an investigation is necessary unless one of two special situations prevails: either (1) postschool learning is determined fully by the prior schooling without any further investment in human capital; or (2) rates of return to schooling and postschool investments in learning are the same. But these are very special situations. Even in an essentially static approach, it becomes necessary to look further into what happens in the postschool years. One of the most debated issues in the economics of education centers around just this problem. Jacob Mincer (1993) has pursued it empirically for the United States over some years, sorting out what part of observed life-earning streams associated with different levels of schooling are in fact attributable to postschool investments, whether in direct outlays or in forgone earnings.

As soon as we shift to talk about change, further questions arise in the interpretation of life-earnings paths constructed from cross-sectional age-earnings distributions. Only if there is no change across cohorts in the forms of those paths will a construct based on age-earn-

ings data at a given time give an unbiased picture of the experiences of any real population cohort. If later cohorts have generally higher earning streams, the cross-section data will understate the increases of earnings over a life span. If there are inter-cohort declines, the steepness of intra-cohort earnings paths will be overstated.

As we should expect, this is one of the spheres in which there has been a relatively active and pragmatic treatment of the three sorts of societal change listed above as they pertain to the economics of education—cyclical shifts, innovative change as it affects demands for skills, and changing relative skill supplies.

Finally it is necessary to treat some critical questions concerning distributions of knowledge and incentives among a population. One of the most important developments in microeconomic theory over the past generation has been the evolution of the economics of information. I have already referred to this indirectly in earlier remarks concerning specialization. But it has much wider and more profound implications for economic theory in so far as that theory is built on one or another concept of “rationality” in human behavior. It calls on us to reassess incentive structures. In the real world what may they imply for the “rationality” of behavior with respect to educational decisions made by individuals? And what about decisions in the use of whatever human capital the individual may have acquired? At this point, where incentives meet motivations, goals, and values, the economist's concerns must interact with the concerns of both psychologists and philosophers. That is a large order. Here I shall cut it down to just two questions, centered primarily, in both cases, on what might be labeled societally “perverse” incentives.

First is the problem of understanding decisions of educational laggards in an affluent society in which schooling is available to all. If the importance of basic general education is so evident, why do many youths remain virtually illiterate, as happens in the United States even today? Does this come back to ignorance of the knowable, or to a lack of economically rational motivation, or to both? For that matter, is a negative educational motivation economically rational in a subsociety that presents some youths with perverse economic incentives? Is the problem one of short time horizons with heavy subjective discounting of potential future returns? If so, why those high discount rates? Or are immediate returns to time spent in criminal activities just too tempting

relative to the risks, ethical values forfeited aside?² These are critical questions not only for sociologists and psychologists, but for economists as well. There is a clear difference between the psychologist and the economist in approaching this problem, however—a difference that dictated use in the first sentence of this section of the word “incentives” rather than “motivations.”

Or, to consider the issue from a different angle, what about the slow reactions of educated people who resist or delay in responding to changes in their future prospects that are relatively easy to predict? Is this in fact an important phenomenon that slows progress for both society and the individual? Or is it bound up with a lack of readiness to take the chances inherent in creative action? Can economists contribute anything here? Perhaps so.

Some further light might be shed on such questions by taking another look at a microeconomic theory of the firm in a world of uncertainty. Over many decades G. L. S. Shackle developed and honed a theory of behavior of the firm in the face of uncertainty (not insurable risk). He introduced the idea of “potential surprise,” favorable or unfavorable, in focusing attention.³ Relatively small variations in likely eventualities would not have such an effect. This proposition may have both a psychological and an economic basis. Economically, the pursuit of information and planning of changed actions or policies are costly, both in direct outlays and in the value of forgone uses of time. Psychologically, there may also be a conservation of effort so long as motivations to avoid potential surprise are not strong. This line of thought brings us to two practical hypotheses with reference to those who drag furthest behind and those who will lead in a changing environment, respectively.

First, looking at educational decisions in this way should help us to understand the seeming irrationality of the disinterest shown in even elementary schooling among members of subpopulations whose entire immediate environments are loaded with anti-education incentives. Not only is the future heavily discounted; in addition any subjective sense of favorable potential surprise associated with schooling is distant. In such considerations, economics and psychology are joined.

Second, the lower the cost of expanding knowledge and the greater the capability for involvement in directed change, the more economically sound and pleasurable will be involvement in innovative actions.

This is where contributions of education to entrepreneurial leadership may come in. Unfortunately, however, incentives can be as perverse among some of the presumably well educated as among the educational laggards—perverse not so much in terms of individual financial incentives as societal benefits and costs.

Conclusions

The bottom line is in the uses of time over and across time. In St. Augustine's remarkable treatment, time is seen to exist only in retrospect (as memory) and prospect (as expectation). The present is no more than a transition from past to future. But past events make the future, and today's future is tomorrow's past. Change is everywhere, even in a society that is repetitive in the turnover of events and the ways in which people make use of time as they move through their life cycles. However, societal change today is much more than repetition as successive cohorts pass through time. Education has played and continues to play a significant part in that process, even as it is also a response.

Conceptually static models of the economics of human resource development and utilization are simplifications that provide a first step toward understanding the decisions that make up economic life. But these first steps can be misleading. Simple repetition in the purest form is in fact an impossibility today, and what can be seen at any given time is already the reflection of relevant recent and prospective changes. Those changes include population growth, shifts in the skill mix of the population (due both to schooling and out-of-school learning), and technological innovations—all of these in both the recent past and in expectations for the future. Any one of them might predict at least directions of change in an otherwise static human investment model, but it is humbling indeed to take all of them together along with changes in the pace and mixes of change. To see this, consider what might happen to human investments and indeed to economic life in general, if the really big change came—a cessation of change! Can we even imagine such a situation in the next generation, given what we see around us and the very nature of human nature?

And so I come to the fundamental generalization—that what we observe is itself a function of change, and so is most of what we may say in general terms concerning the economics of education. Despite changes that will come (for better or for worse) in the educational structures and contents of a future world, this much we can say with assurance: education itself contributes to change. And a world of change calls for learning both within and outside of schools. It calls for general education as a preparation for future learning, for specializations that can cope with change, and for both applied and theoretical learning.

The empirical referents in even our static models of human investment decisions and benefit-cost theory are built on expectations concerning a reality that is always changing. That reality is embedded in the flow of time. And so it happens that we are facing and affecting change, whether we see things that way or not. In the long run only a conscious awareness of this fact can bring us closer to understanding the events and the problems that surround us and in which we are inevitably enmeshed.

Notice, however, that none of the relatively firm pronouncements just set forth says anything about the underlying purposes of education, nor do they take note of the origins of economics in moral philosophy, so wisely stressed by Harry Johnson (1972) in his commentary at a conference on “The Equity Efficiency Quandary in Education.” Yet the present paper was written to communicate with an essentially academic audience from diverse disciplines, and on the same day I presented to a group of economists a paper that I called “The Day Aristotle Visited an American School System.” Both Harry Johnson in the late twentieth century A.D. and Aristotle in the fourth century B.C. stress the importance of reason, and both challenge us to look beyond narrow boundaries in our thinking to ask what is really important: in the present context, education for what? In Aristotle, this is lifelong learning, from early upbringing of children (to “moral virtue” by inculcation of good habits), on to an unending pursuit of wisdom, both theoretical and practical. In Harry Johnson it is essentially the same, though he too is a man of his own time. The twenty-first century will soon be here. Perhaps it will call us to seek a wisdom less bounded by formal academic disciplines and more alert to the human questions that are already calling upon us to probe the roots of education in our day.

NOTES

1. For early examples, see Chaudhri 1968, and especially, Welch 1970.
2. Richard Freeman (1992) has addressed this problem recently in a working paper for the National Bureau of Economic Research.
3. For an application of some of his ideas to investments in human beings, see Bowman 1972.

References

- Becker, Gary S., and Kevin Murphy. 1990. "The Division of Labor, Specialization and Economic Progress." Department of Economics, University of Chicago.
- Bowman, Mary Jean. 1072. "Expectations, Uncertainty and Investment in Human Beings." In *Uncertainty and Expectations in Economics*. Essays in Honor of G. L. S. Shackle. Edited by C. F. Carter and L. L. Ford. Oxford: Blackwell.
- Chaudhri, D. P. 1968. "Education and Agricultural Productivity in India." Ph.D. dissertation, University of Delhi.
- Freeman, Richard. 1992. "Crime and the Employment of Disadvantaged Youth." Working Paper No. 3875. National Bureau of Economic Research.
- Johnson, Harry G. 1972. "The Alternatives Before Us," *Journal of Political Economy* 80, 3, part 2, (May-June). Investment in Education: The Equity-Efficiency Quandary, S280-9.
- Mincer, Jacob. 1993. In *Studies of Human Capital: Collected Essays*. Jacob Mincer, ed. London: Edward Elgar.
- Rosen, Sherwin. 1983. "Specialization and Human Capital," *Journal of Labor Economics* 1: 43-49.
- Schultz, Theodore, W. 1975. "The Value of the Ability to Deal with Disequilibria," *Journal of Economic Literature* 13, 3 (September): 827-46.
- _____. 1990. *Restoring Economic Equilibrium: Human Capital in the Modernizing Economy*. Cambridge, MA: Basil Blackwell.
- Welch, Finis. "Education in Production," *Journal of Political Economy* 78 (January-February): 35-39.

INDEX

- Abraham, Ansley A., Jr., 16
- Accelerated schools
for at-risk students, 25
benefits of, 3
performance of, 29-30
recommendations for development, 2
strategy to create, 25-27
transformation to, 28-29
- Accountability
in accelerated schools concept, 26-27
demands for educational performance, 75
test scores as instrument of, 79-80
- Achievement tests, standardized, 77-78
- ACT scores, 79
- Adaptation, 166-67
- Alexander, Karl, 112
- Alsalam, Nabeel, 112, 130
- Alwin, Duane, 112
- "America 2000," 75
- Armor, D., 62
- Astin, Alexander, 112, 118
- At-risk students
accelerated schools for, 25
assets and strengths of, 27
defined, 2, 11
economic analyses of investments in, 18-23
economic impact of, 14-18
effective approach to educating, 24-25
effect of compensatory programs for, 24
proportion of total school population, 12-14
understanding incentives of, 170-72
- Barnett, W. Steven, 21, 22
- Basic Competency Test (BCT),
Alabama, 4, 55-58
- Becker, Gary, 168
- Berlin, Gordon, 17
- Berrueta-Clement, John R., 22
- Birnbaum, Max, 115
- Card, David, 22
- Cass, James, 115
- Catterall, James S., 21
- Celis, William III, 54
- Change, societal, 164-66, 170, 172
- Chubb, John E., 30, 61
- Clune, William H., 77
- Cohen, Judith, 53, 54
- Coleman, James S., 30
- Coleman Report
findings and contribution of, 40-42
interpretation of findings, 46-47
- Conaty, Joseph, 112
- Congressional Budget Office, 13, 38
- Consumer Price Index, 115
- Daniel Webster School, San Francisco, 29
- Data normalization, 150-51
- Data sources
for analysis of resource allocation in higher education, 146-50
for analysis of school accountability, 93
for college choice and higher earnings analysis, 112, 114-15
for trends in higher education, 140-41
- Dayton, Charles, 19
- Disequilibria concept, 165-66
- Dominitz, Jeff, 78, 97, 98, 99
- Dossey, John A., 94t
- Dropout prevention programs, 18-21
- Dropouts. *See* At-risk students
- Duncan, Greg J., 125
- Dyer, Henry S., 81, 94

Earnings

- dependence on higher educational experience, 129-30
- for males choosing observable college characteristics, 122-23t, 124-26
- as output of higher education, 113-14

Earnings, lifetime (dropouts), 20**Education**

- contribution to change, 173
- definition of, 163
- economics of, 164, 173
- preparation in early years for, 10, 166-67
- purposes of, 173
- societal change in economics of, 164-66, 170, 172-73

Education, accelerated. *See* Accelerated schools**Education, general, 9-10**

- to deal with change, specialization and disequilibria, 166-67, 173
- defined, 167-68
- levels of, 168

Education, higher

- cost sharing of, 155-57
- forces impacting, 141-43
- information sources on trends in, 140-41
- outputs of, 113
- quality of, 15-16
- trends or cycles in, 140-44

Education, remedial

- in colleges and universities, 16
- current system, 2
- effect on at-risk and disadvantaged students, 24
- recommendations for effective, 24-25
- See also* Accelerated schools

Educational performance indicators. *See* Performance indicators**Educational process**

- for at-risk students, 24
- output related to input, 41

Efficiency, 48-49**Entrepreneurship, 165-66*****Equality of Educational Opportunity. *See* Coleman Report*****Equity concept, 3, 35, 36, 48-49****Expenditure, higher education**

- components of, 145-50
- data normalization, 150-51
- trends or cycles in, 151-55

Expenditure, school

- determinants of instructional, 42
- effect of policy based on differences in, 49-53
- interpretation of, 48-49
- relation to student performance, 3-4, 37-40, 42-46, 53-63

Friedman, Milton, 30**Fullerton, D. J., 52****Gain indicators**

- as descriptive indicator, 99
- measurement by, 81-82, 93-94
- relation to value-added indicators, 95
- to report gain in educational level, 6

Garibaldi, Antoine, 16**Goldman, Jay P., 79****Greeley, Andrew, 30****Griffin, Larry, 112****Hansen, W. Lee, 159`****Hanushek, Eric A., 42, 44t, 61, 79, 94, 112****Hauptman, Arthur M., 139****Haveman, Robert, 18, 23****Henderson, James M., 98****Higher Education General Information Survey (HEGIS), 115****Hirschman, A. O., 140****Hoffer, Thomas, 30****Hollibrook Elementary School, Houston, 29****Human capital**

- incentives to specialize from returns to, 168
- as mainspring of growth, 165-66
- postschool investment in, 169

- Illiteracy, 15
- Information
   cost to pursue, 171
   theory of, 170
- Inputs, educational process, 41-42
- Investment, educational
   assessment of dropout prevention programs, 18-21
   in at-risk students, 28-30
   in early years, 166-67
   as investment in learning, 163
   at preschool and higher education levels, 21-22
   preschool benefit-cost analysis, 21-22
   societal changes in, 164-65
- James, Estelle, 112, 116
- Jamison, Dean T., 42
- Johnson, Harry, 173
- Jung, Richard K., 14
- Kelley, Carolyn, 23
- Kennedy, Mary M., 14
- Knapp, Michael S., 25
- Koretz, Daniel, 13
- Kozol, Jonathan, 36, 58, 59
- Krueger, Alan, 22
- Learning, postschool, 169-70
- Level indicators
   in evaluate school performance, 95-96
   questioned reliability of, 78-79
   See also Gain indicators
- Levin, Henry M., 14, 15, 20, 23, 24, 25, 30
- Linn, Robert L., 81, 94
- McCarthy, Jane, 29
- McDill, Edward L., 13
- McKnight, Curtis C., 39
- McNett, Ian, 13
- McPherson, Michael S., 139
- Major subject returns, 130-31
- Masten, Charles L., 22
- Meyer, Robert H., 78, 97, 98, 99
- Miller, Herman, 111, 124
- Mincer, Jacob, 169
- Moe, Terry, 30
- Morgan, James N., 125
- Muller, Thomas, 13
- Murnane, Richard J., 62, 94
- Murphy, Kevin, 168
- National Academy of Sciences, 15
- National Assessment of Educational Progress, 15, 79, 93
- National Center for Educational Statistics, 13, 14
- National Commission on Excellence in Education, 23
- National Longitudinal Study of the High School Class of 1972 (NLS-72), 112, 114
- Nation at Risk. See* National Commission on Excellence in Education
- Natriello, Gary, 13
- Neuberger, Egon, 116
- Oates, Wallace, 52
- O'Day, Jennifer, 12, 15, 77
- Orland, Martin E., 14
- Outputs, educational process
   in higher education, 113
   measurement of, 41-42
- Owens, John D., 17
- Paik, Il-Woo, 19
- Pallas, Aaron M., 13
- Parents' role
   in accelerated school concept, 27-28
   in remedial education, 24
- Patton, Michael J., 81, 94
- Performance, student
   distribution of, 41
   relation to school expenditure, 3-4, 37-40, 42-46
- Performance, teacher, 47-48
- Performance-based tests, 78

Performance indicators

- to assess school performance, 75-76
- effect of valid and invalid, 95-96
- as measure of school contribution, 79
- See also* Gain indicators; Level indicators; Production function approach; Value-added indicators

Performance indicator system, ideal, 97-98

Perry Preschool Project, 21-22

Peterson, John M., 25

Postsecondary Education Transcript Study (PETS), 112, 115

Powell, Brian, 79

Production function approach, 40-42

- See also* Inputs, educational process; Outputs, educational process

Public policy

- impact on trends in higher education, 142

- promotion of performance-based, 60-62

Public service cost, 17-18

Quandt, Richard E., 98

Reed, Ritchie, 111, 124

Rivkin, Steven G., 42

Robledo, Maria del Refugio, 20

Rosen, Harvey S., 52

Rosen, Sherwin, 168

Rumberger, Russell, 12, 15

St. John, Edward P., 22

San Antonio Independent School District v. Rodriguez, 35-36

SAT scores

- as index of institutional selectivity, 117
- as level indicators, 79
- as measure of student performance, 38-39

Schlesinger, Arthur M., Jr., 140

School report cards, 75

Schools, public

- financing of, 35
- pupil-teacher ratios, 40

Schultz, T. W., 165, 166

Serrano v. Priest, 36

Shackle, G. L. S., 171

Shapiro, Morton Owen, 139

Shields, Patrick M., 25

Smith, Marshall, 12, 15, 77

Solmon, Lewis, 111, 124

Specialization

- effect of, 166
- general education with, 167-68, 173
- incentives and requirement for, 168-69

Stanford Accelerated Schools Project, 2-3, 25

Steelman, Lala Carr, 79

Stern, David, 19

Still, Suzanne, 29

Sum, Andrew, 17

Summers, Anita, 126

Taylor, Lori, 79

Teachers

- differences in performance, 47-48
- perceived role in traditional schools, 26
- public school, 40
- in remedial education, 2, 24
- role in accelerated school concept, 27-28

Testing, annual, 81-82, 99-100

Tests, standardized, 78-80

- See also* Basic Competency Test (BCT), Alabama; Level indicators

Test scores

- failure to measure school performance, 76, 79-81, 97, 99
- simulations to assess, 82-92, 103-8
- See also* ACT scores; SAT scores

Tiebout, Charles M., 52

To, Duc-Le, 112

Turnbull, Brenda J., 25

- U. S. Department of Commerce, Bureau of the Census, 12
- U. S. Department of Education, 75
- Value-added indicators
  - factors determining quality of, 99
  - as measure of learning, 6
  - relation to gain indicators, 95
- Ventresca, Marc, 13
- Wachtel, Paul, 111, 124
- Wainer, Howard, 79
- Wales, Terence, 111
- Weisberg, Alan, 19
- Weisbrod, Burton A., 18
- Wendling, Wayne, 52, 53, 54
- Willms, J. Douglas, 30
- Wise, David, 129
- Wolfe, Barbara, 126
- Wolfe, Roberta, 18, 23

About the Institute

The W.E. Upjohn Institute for Employment Research is a nonprofit research organization devoted to finding and promoting solutions to employment-related problems at the national, state, and local level. It is an activity of the W.E. Upjohn Unemployment Trustee Corporation, which was established in 1932 to administer a fund set aside by the late Dr. W.E. Upjohn, founder of The Upjohn Company, to seek ways to counteract the loss of employment income during economic downturns.

The Institute is funded largely by income from the W.E. Upjohn Unemployment Trust, supplemented by outside grants, contracts, and sales of publications. Activities of the Institute are comprised of the following elements: (1) a research program conducted by a resident staff of professional social scientists; (2) a competitive grant program, which expands and complements the internal research program by providing financial support to researchers outside the Institute; (3) a publications program, which provides the major vehicle for the dissemination of research by staff and grantees, as well as other selected work in the field; and (4) an Employment Management Services division, which manages most of the publicly funded employment and training programs in the local area.

The broad objectives of the Institute's research, grant, and publication programs are to: (1) promote scholarship and experimentation on issues of public and private employment and unemployment policy; and (2) make knowledge and scholarship relevant and useful to policymakers in their pursuit of solutions to employment and unemployment problems.

Current areas of concentration for these programs include: causes, consequences, and measures to alleviate unemployment; social insurance and income maintenance programs; compensation; workforce quality; work arrangements; family labor issues; labor-management relations; and regional economic development and local labor markets.