Volume Title: The Measurement of Economic and Social Performance

Volume Author/Editor: Milton Moss, ed.

Volume Publisher: NBER

Volume ISBN: 0-870-14259-3

Volume URL: http://www.nber.org/books/moss73-1

Publication Date: 1973

Chapter Title: Measuring Performance in Education

Chapter Author: Alice M. Rivlin

Chapter URL: http://www.nber.org/chapters/c3619

Chapter pages in book: (p. 411 - 437)

Measuring Performance in Education

ALICE M. RIVLIN

THE BROOKINGS INSTITUTION

THIS is supposed to be a practical conference dealing with national accounts and how to make them more useful. An early communication from the planning committee stated that the purpose of the conference was "to provide specific suggestions for changes in the existing accounts with an eye toward making the accounts more useful as measures of economic and social performance as well as increasing their analytical usefulness." In this context, my specific assignment is to discuss the measurement of performance in education.

It seems to me important to take a broad view of "national accounts" and to recognize that the present national income and product accounts are only one set of a large number of possible accounts which would be useful for various purposes. There is no point in plunging into endless wrangles about what should be included in GNP or national income (should nonmarket activities be included? what about the use of leisure time?). While some changes in the national income and product accounts may be in order, the really interesting questions relate, not to changing the present accounts, but to supplementing them. The problem is not so much to expose what is wrong with the GNP, but to decide what *other* information we need to collect on a regular and sustained basis, and how this new information can be linked to that already in the national accounts.

Similarly, it seems to me useful to take a broad view of "education" and allow for subcategories within it. Broadly speaking, education is "imparting skill and knowledge." Many different activities could come under this general heading, and there is no point in fussing over exactly what should be included (is television education? is on-the-job training education?). For some purposes, it is useful to focus on regular class-

NOTE: The views expressed in this article are those of the author and do not represent the views of other staff members, officers, or trustees of The Brookings Institution. The author is indebted to Edward Denison, Charles Schultze, and Robert Hartman for comments on an earlier draft.

room instruction. For others, it is useful to include other activities that produce skill and knowledge. What is needed is a detailed and flexible series of accounts that will allow the user to add up the various boxes as he sees fit.

OBJECTIVES OF NATIONAL ACCOUNTS

One would hope that a good set of national accounts would help to answer three types of questions: where is it going; how are we doing; and how can we do better.

Where-is-it-going questions. It is clearly useful—if one wants to understand the economy—to know where the national output is coming from and what is happening to it. How is total output allocated between consumption and investment, or among the factors of production? How much is being produced in various sectors of the economy? The income and product accounts provide a basis for examining the allocation of production and resources in the market sector of the economy. For many purposes, it is also useful to know how nonmarket production is allocated and how the nation uses its time.

Since education is now recognized not only as a major economic activity, but also as an important form of investment, one might hope for subtotals in the national accounts that would give a sense of the relative importance of education in the national productive effort. What proportion of national output is devoted to education? How much investment takes the form of education? How much education is produced in the business or government or household sectors? How much time is used for education?

How-are-we-doing questions. The second set of questions on which national accounts should shed light are those of particular interest to this conference—how the system is performing over time. It is clearly useful to correct for price changes in order to find out if total output is rising and how fast. It is useful to relate the outputs to the inputs—to measure productivity—and to determine how fast the nation is adding to its stock of capital. One might hope, in this connection, that national accounts would tell us something about performance in education. Is the real output of the education industry rising? Is it rising faster than the inputs? How rapidly is educational activity adding to the stock of human capital?

How-can-we-do-better questions. The most important use of national accounts is as input to models designed to explain how the economy works and to provide clues to improving its performance. The most obvious use is to discover relationships between aggregates, such as total

income, investment, consumption, or employment. One might, however, want to use national accounts data as input to disaggregated models designed, for example, to throw light on the way the education sector interacts with the rest of the economy. One might hope to find in the national accounts information that would be useful in answering such questions as: Would national output be increased by devoting more resources (or less) to education or by shifting resources from one kind of education to another? How would shifts in the education sector affect employment or productivity or price levels?

This paper will focus on these three sets of questions in turn, examine the extent to which they can be answered by the existing income and product accounts, and explore what else might be useful.

WHERE-IS-IT-GOING QUESTIONS

Education—whether broadly or narrowly defined—is a challenge to national accountants. It is a service, not a good; it is partly consumption and partly investment; the consumption is partly immediate, and partly spread over time like consumer durables; it is produced by business enterprise and by households, as well as by government; it is partly produced outside the market altogether. At best, it takes a lot of arbitrary decisions to devise an accounting system to answer even the simplest where-is-it-going questions: How much of national product is devoted to education? How much of national income originates in the education industry?

The first thing one notices in examining the present income and product accounts to see how they treat education is that "education" is hard to find. It would be useful—now that there is such interest in the education sector—for the government to publish some regular tables breaking out both expenditures for education and income originating in the "education industry."

If one tries to do this from the currently published statistics, one encounters some anomalies of the national accounts; for example, the different treatment of public and private educational institutions that would seem to the casual observer to be doing exactly the same thing.

Suppose that a private university becomes a branch of the state university system. Even if there is no change in enrollment or expenditures, there will be changes in the national accounts. In the national income account, income originating in education will drop because interest paid by the government is excluded from the national accounts. The rationale for excluding government interest payments "stems, as a practical matter, from the fact that the bulk of government debt was created to finance

TABLE 1

Enrollment in Elementary and Secondary Schools and in Institutions of Higher Education, Public and Private, 1950, 1970, and 1980

	Enrollment (millions)		Percentage Public
	Public	Private	
Elementary and secondary education			
1950	24.0	3.2	88.2
1970	45.6	5.7	88.9
1980 (projected)	45,6	5.4	89.4
Higher education			
1950	1.4	1.3	51.9
1970	5.3	2.0	72.6
1980 (projected)	8.7	2.4	78.4

SOURCE: U.S. Office of Education, *Digest of Educational Statistics, 1969,* 1970, 1950 data, Table 3, p. 3; U.S. Office of Education, *Projections of Educational Statistics to 1979–80,* 1971, 1970 and 1980 data, Table 3, p. 20, and Table 12, p. 29.

wars and current expenditures."[1] If one is primarily interested in the federal level, this rationale carries some weight, but at the state and local level, where almost all public education occurs, wars are not fought and debt financing bears a much closer relationship to capital outlays.

The magnitudes involved are not huge, but neither are they entirely insignificant. For example, for public elementary and secondary schools in 1968–69, interest on school debt was about $1.0 billion out of total expenditures of $35.5 billion.[2]

In recent decades, enrollment in public education has grown considerably faster than in private education, especially at the college and university level. As education has become more public, the relative importance of the excluded costs has increased and the growth of education relative to other sectors of the economy has been correspondingly understated. The projections shown in Table 1 anticipate continued growth in the proportion of students in public institutions. On the other

[1] U.S. Department of Commerce, *National Income,* 1954 edition (A Supplement to the *Survey of Current Business*), p. 35.

[2] U.S. Department of Health, Education and Welfare, *Digest of Educational Statistics,* 1969 edition, p. 52.

hand, new developments such as widespread adoption of Friedman-type voucher plans might shift large numbers of students from the public to the private sector without any change in the resources devoted to education. It clearly makes little sense to treat the costs of public and private educational institutions differently in the national accounts.

Some other possible real world events suggest some more serious questions about the completeness of the present accounts. Suppose, for example, that a corporation (or a group of corporations) induces the government to set up a public vocational school which reduces the corporation's need to engage in on-the-job training. Trainees, who were being paid—partly to work and partly to learn—now become unpaid students.

In the national accounts, such an event would show up as a decrease in factor payments by the industry in question and an increase in resources devoted to education. The increase in education, however, is likely to be less than the drop in industry factor payments because the earnings of trainees show up in the national accounts while the foregone earnings of students do not. The potential for shifts from vocational schooling to on-the-job training or vice versa is quite large and it would be useful to be able to see more clearly in the accounts what shifts are occurring.

Or suppose a group of mothers starts a cooperative nursery school, charging tuition and paying themselves salaries as part-time teachers. This event would increase the income originating in education as reported in the national accounts since the educational services of mothers show up in the accounts only if they are paid salaries. The actual amount of time the mothers spent "educating" their children, however, might not increase, indeed is likely to decrease.

The current potential for this kind of shift is very large. In 1968, only 4 million of the 12 million three- to five-year-olds in the country were enrolled in school. The current push for day care and preschool could easily put an additional 4 million children into formal programs partially financed by the government and largely staffed by their mothers. An expenditure of $1,500 per child (cheap for a full-day educational program) would increase reported educational outlays by $6 billion, perhaps half or more of which would involve substitution of paid for unpaid instruction by the same mothers.

Finally, suppose a cable television network starts a "university of the air" with a broad curriculum of courses. Some people who would ordinarily have attended regular schools and colleges begin taking these

courses instead. (A procedure may be established for awarding them degrees by examination.) In this case, factor costs of education reported in the national accounts would fall and factor payments in the radio and television industry would rise, unless education became the principal activity of the establishments in the network.

If one were looking at the expenditure side of the accounts, an even odder thing would happen. The expenditures for education would fall, but the compensating increase would be spread over other products, because commercial television is treated as an intermediate product which only shows up in the prices of products made by enterprises that purchase advertising.

The potential for substituting wide varieties of nontraditional education (home courses on cable television or TV cassettes, schools without walls, external degrees, etc.) for traditional education appears to be very great, given the advance of technology and the general dissatisfaction with traditional education. The result could be a vast increase in the amount of resources, including leisure, devoted to genuine learning that would not show up in the national accounts as presently constituted.

The treatment of these rather common events in the national accounts are often regarded as "yes" or "no" decisions involving either difficult distinctions between what is or is not education (when is television "educational"?) or difficult decisions about the inclusion or exclusion of nonmarket costs (student time, mothers' services) or both. Those arguing against inclusion often invoke a *reductio ad absurdum* argument, alleging that the inclusion of a particular item or a particular cost, while logical for the purpose at hand, would necessarily involve—for consistency's sake—a whole series of decisions of highly questionable utility. Inclusion of mothers' time spent educating preschoolers or students' time would involve including other kinds of unpaid activities in the GNP (do-it-yourself hobbies and volunteer work), would blur the distinction between work and leisure, and might even lead to the inclusion of leisure time in the national accounts. Similarly, they argue that including educational TV in "education" would lead to including all TV and radio and books and magazines and records and all kinds of other activities that are not really educational—once one leaves the classroom there is no stopping.

But these arguments reflect the notions that (1) there must be unique classifications for each type of activity, and (2) the parts must add to a single number that accurately reflects fluctuations in economic activity over time, presumably the GNP. These notions were appropriate to an

era of meager statistical information and limited computer capacity when it was hard enough to set up *one* reasonable set of classifications and fill in the boxes, and impossible to allow for multiple classifications and the possibility of adding the same item into the accounts in several alternative ways at the whim of the user. The assumptions were also appropriate when the accounts were primarily used as a measure of the fluctuation of aggregate economic activity in the market sector. Imputations and allowances for nonmarket activity just fuzzed up the numbers and masked the fluctuations of real interest to the users.

But in the 1970s, these assumptions need no longer hold. Statistical information about the economy is more plentiful and still increasing. Computers are easily able to handle multiple classifications, to show users the consequences of classifying the same activity in different ways, even to warn the user when he is double-counting or being inconsistent. Moreover, interest in national accounts no longer centers so heavily on the business cycle and aggregate activity in the market sector, but now arises partly out of interest in the way total resources are allocated among more specific market and nonmarket activities. It seems to be time for a more flexible and detailed system of accounts.

The new accounting framework proposed by the Ruggles is a step in the right direction.[3] The Ruggles divide the economy into three sectors: enterprise, households, and government. "Enterprise" includes nonprofit institutions. A consolidated income and product account would be constructed for the three sectors together and separate income and outlay accounts for each of the three sectors. Capital consumption and capital formation would be estimated for each of the three sectors in comparable terms. Moreover, "development expenditures" (education, training, research) would be estimated for each sector and shown as capital formation. To complement this change, imputed income from past development expenditures would be broken out. Subtotals for "education" could easily be shown in the income and product account (income originating in the education industry, capital consumption in the education industry) and in the income and outlay accounts for the three sectors (development expenditures on education by households, government, and enterprises; and imputed income from past development expenditures in education).

Although it would be difficult to make many of these estimates, the Ruggles system appears to provide a better framework for answering

[3] Nancy Ruggles and Richard Ruggles, *The Design of Economic Accounts*, New York, National Bureau of Economic Research, 1970.

where-is-it-going questions involving education in the economy than present accounts. It would provide a continuous series of estimates answering the questions: What part of national income originates in the education industry? How large is capital consumption in the education industry? What part of national product is devoted to capital formation in the form of education outlays by government, households, or enterprises?

The problem of defining "education" would remain. A flexible series of definitions would be possible—user's choice—under which, say, educational television or on-the-job training could be defined in or out of "education" as the interest of the user dictated. Similarly, nonmarket costs of education, such as mothers' and students' time, which the Ruggles do not include, could be estimated but included in totals only when it seemed useful to do so.

HOW-ARE-WE-DOING QUESTIONS

A one-shot or still-picture set of national accounts is useful, but even greater interest clearly attaches to changes over time, especially changes in real output.

In no sector of the economy is it easy to obtain a measure of real output. In the goods-producing sector where physical units can be counted the problem may seem easy at first glance, but becomes exceedingly difficult as soon as one begins to grapple with quality changes. The quality of almost all goods produced in the economy changes somewhat over time and for many goods the changes are rapid, continuous, and in several dimensions. Price deflators of goods as variable as toys or books or musical instruments are obviously extremely arbitrary. Measuring the output of services is also difficult, not only because quality varies, but because the unit of service is sometimes not obvious, and because, as Fuchs has pointed out, the customer himself often plays an important role in determining the quality of the service.[4]

In the case of government services which are given, not sold, to the public, the national accountant is at an even greater loss. The market does not place a value on these goods—there is no market price to deflate—so costs must be used as a measure of value. The tradition of national accounting is that costs are deflated by some measure of price changes in the input factors (wages, materials) to obtain measures of real output. Changes in productivity of the inputs are ruled out by definition.

[4] Victor Fuchs, assisted by Irving Leveson, *The Service Economy*, New York, National Bureau of Economic Research, 1968, p. 194.

In deflating the U.S. national accounts, education outlays are divided into three parts: payments to business, payments for construction, and payments to employees. Payments to business are deflated by using the most appropriate component of the wholesale price index. Payments for construction are deflated by using the Commerce Department's "building materials and construction index," adjusted for changing profit margins. Payments to employees—by far the largest education cost item—are deflated by an index of average wages in education. No attempt is made to adjust for shifts in the mix of employees—janitors, second grade teachers, and college presidents count the same—or for changes in the "quality" of specific types of employees. As Schultze points out, this highly aggregative approach could hide important shifts occurring within the education sector.[5] If colleges are expanding faster than elementary schools, average wages in education are likely to be rising because professors are paid more than elementary school teachers. A professor, however, represents "more educational input" than a second grade teacher, as valued by the market, in the same sense that a Cadillac represents "more car" than a Volkswagen. In this instance, deflating education costs by an aggregate wage index would tend to underestimate the shift in education input that was occurring. In principle, therefore, the national accounts could be improved by disaggregating education employment into more homogeneous categories, although in practice the adjustment may not make much difference.

No amount of disaggregation of inputs, however, will provide a basis for answering the how-are-we-doing questions in the education sector. As long as cost is used as a proxy for value there is no way to compare inputs with outputs or to see whether a given "amount" of education is being produced with fewer resources. It does not tell us anything about the impact of the education on the recipients and whether they are receiving more or less education than students in an earlier period. So the question arises: Are there better ways to measure output or performance in education?

Before this question is addressed, it should be noted that the difficulties of measuring the output of collective or indivisible goods, discussed at length in Mancur Olson's paper, do not seem especially important in education. There are two kinds of divisibility: divisibility of the good or service itself and divisibility of the benefits flowing from that good or service. The service "education" is not hard to divide into units—years,

[5] Charles L. Schultze, Comment on Mancur Olson, "Evaluating Performance in the Public Sector," in this volume.

hours, etc.—nor is it difficult to exclude potential customers from the acquisition of these units of education. Students can be excluded from Princeton or any other university by law, price, or entrance requirements. The private benefits that accrue to individuals from the acquisition of these units of education (better jobs, higher incomes, etc.) are likewise divisible. There may also, of course, be benefits of education that spill over to people who did not receive the education themselves. If education reduces crime, then the benefits of education go not only to the potential criminals who receive the education but to their potential victims as well. The size of these collective benefits relative to the private ones bears on the question of how education should be financed—to what extent potential victims should be taxed for subsidizing the education of potential criminals, for example—but the existence of collective benefits does not keep us from measuring how much of this service called education is actually being produced.

There are two principal reasons why measuring output in education is so difficult—both of which would remain if it were produced in the private market. First, the quality of education varies greatly and in many different dimensions. Second, the customer (the student) helps produce the education and the quality of the output is hard to disentangle from the customer's input.

It is not hard to think of plausible measures of output in education—the problem is choosing among them. At least six types of measures come readily to mind.

1. Body Counts

As a start, one might consider simply counting the number of customers served and using enrollment as a measure of output in education. But the customers are served in such widely different ways by various kinds and levels of education that one can only conclude there must be a better way of measuring output than simply counting the bodies passing through the system.

2. Student Exposure Measures

Part of the difficulty with body counts is that enrollees are exposed to varying amounts of education: The problem is particularly obvious in higher education where part-time students abound and even supposedly "full-time" students take different course loads. This problem can be alleviated by using a unit of student exposure, such as the credit hour, as a measure of output. This is the approach used by June O'Neill in a

recent study of inputs and outputs in higher education.[6] One could conceive of extending O'Neill's approach by developing measures of student exposure at other levels of education. In elementary and secondary education, almost all students study "full-time" and for approximately the same number of hours a day, although the length of the school year varies. "Student days" might be a suitable exposure measure for grades 1 through 12. For preschool and vocational programs, one would have to adjust for varying school hours as well as varying numbers of days of exposure.

The trouble with exposure measures, of course, is that the hours or days to which students are exposed may differ greatly in quality and intensity. Students are served in larger or smaller groups (is a credit-hour in a class of 10 more valuable than a credit-hour in a class of 100?) and by teachers of varying skills and training, working with varying amounts of equipment and ancillary resources. One might, therefore, decide to derive a measure of "quality" or intensity of education which could be used to weight the exposure units. If one uses inputs as a measure of intensity of exposure, however, one is forced to derive weights for the importance of the various inputs. There is no obvious way of doing this except to use either their price or their cost. In education, price has no clear meaning because of the large subsidies, but one might construct a measure of educational output which weighted students (or credit-hours) by some measure of relative cost.

A weighted output measure in, say, higher education might classify students (or credit-hours) by the type of institution attended. The cost per student in each type would serve as weights on the assumption that relative cost roughly measures relative value.[7]

3. Student Attainment Measures

Exposure measures (whether or not weighted by cost) are still indexes of student time and other inputs to the education process, not of what the process produces or achieves. It would be preferable to have a true output measure. One possibility would be to accept the education system's own definition of output, and assume that a student who has

[6] June O'Neill, *Resource Use in Higher Education: Trends in Outputs and Inputs, 1930 to 1967*, Berkeley, Carnegie Commission on Higher Education, 1970, Chap. 2.

[7] For an attempt to do this, see *ibid.*, pp. 12–15 and 44–47, in which O'Neill adjusted student credit hours to reflect changes in levels of instruction (lower division, upper division, graduate) and type of institution, but found that the adjustment made little difference.

completed one level of education or has been promoted to the next level has given evidence of some kind of educational success.

Some have suggested using the proportion of students moving from one grade level to the next as a measure of output (or the proportion of students "left back" as a measure of nonsuccess). In the same spirit one might use high school graduations or college degrees as proxies for educational output.

In a country which had an elitist school system with uniformly high standards for promotion and graduation these measures might be defensible. In the United States, however, educational standards vary from high to nonexistent, and there is no reason to think that completions or graduates convey more information about output than body counts of enrollees. That the two move together may be seen in the comparison of degrees and enrollment in higher education in the 1960s presented in Table 2.

If one wants to take student intellectual achievement as the output of education, then —in the United States, at least—it will probably be more accurate to measure that achievement directly, rather than to use degrees or completions as proxies for it.

TABLE 2

A Comparison of Enrollment and Degrees in Higher Education, 1959–70

Academic Year	Enrollment (millions)	Degrees (millions)	Degrees as Percentage of Enrollment
1959–60	2.636	.389	15
1960–61	2.876	.395	14
1961–62	2.956	.414	14
1962–63	3.163	.444	15
1963–64	3.406	.494	15
1964–65	3.699	.530	14
1965–66	4.066	.551	14
1966–67	4.302	.591	14
1967–68	4.572	.667	15
1968–69	4.812	.764	16
1969–70	5.014	.784	15

SOURCE: U.S. Office of Education, *Projections of Educational Statistics to 1979–80*, 1971, Table 19, p. 36, and Table 21, p. 42.

4. Student Achievement Measures

If the purpose of education is to impart skills and knowledge to students, the most direct way to measure output is to measure those skills and knowledge. This reasoning suggests the use of test scores as proxies for output in education.

This is the age of testing. Considerable effort has gone into devising and standardizing a wide variety of tests of intellectual skills and accumulated knowledge. Billions of man-hours of student and teacher time are devoted to taking, administering, grading, analyzing, and discussing standardized tests. One might hope that all the effort would tell us something about output or performance in education.

Remarkably, almost no information presently exists which would give a basis for constructing an index of change in educational test scores over time. After shrill and sustained resistance to the idea from the school establishment, a "national assessment of education" is being undertaken for the first time this year. A battery of tests designed to reveal intellectual skills and knowledge of specific areas (science, music, etc.) is being given to a large sample of children and adults. Results are just beginning to appear. If this effort is repeated, periodically in the future, there will be a basis for comparisons over time. At present, however, one would be hard put to construct an index. Some of the test-making companies have information over several years from the samples they use for standardizing.[8] A few cities have administered the same set of tests for several years, but national information about changes over time is almost nonexistent.

Two major national studies, however, have given extensive test score information for cross-sections of students. Project TALENT administered a big battery of tests to a sample of about 100,000 high school students in 1960 and collected a lot of other information about these students and the schools they attended.[9] Another survey, made pursuant to the civil rights act of 1964 and usually known as the Coleman study, tested an even larger sample of children (over 600,000) both at the elementary and high school levels and collected data on their schools.[10] These studies (and a few others in which tests were administered in a single city or

[8] U.S. Department of Health, Education and Welfare, *Toward a Social Report*, Washington, D.C., 1969, p. 67.

[9] For a discussion of the study, see John C. Flanagan et al., *Designing the Study*, Project TALENT, Monograph No. 1, Pittsburgh, University of Pittsburgh, 1960.

[10] James S. Coleman et al., *Equality of Educational Opportunity*, Washington, D.C., 1966.

other geographic area) have provided data for a series of attempts by statisticians to relate test scores (viewed as measures of outputs) to school characteristics (teacher-student ratios; age, training, or verbal aptitude of teachers; size of school; age of building; expenditures per student; etc.) and characteristics of students and their families (race, socioeconomic status, etc.).

The most general result of these statistical studies has been the finding that variables reflecting the socioeconomic characteristics of students and their families explain most of the variation in test scores, and variables reflecting school characteristics or resource inputs explain very little.[11]

These results should not be exaggerated—they do not prove that "schools don't matter"—but they certainly provide a basis for considerable skepticism about using test scores as measures of the output of the education industry as such. Test score changes may primarily reflect changes in the school population and the way it is mixed, rather than the productivity of school resources themselves. Some investigators have tried to control for the student inputs by explaining "achievement," given "ability." But serious challenges have been made to the independence of IQ tests (the usual "ability" measure) of school influences.

There are other objections to the use of standardized tests as measures of the output of education. There is the difficulty of establishing weights for different kinds of tested skills (should reading comprehension be weighted more heavily than mathematical facility?). This problem occurs in constructing tests as well as weighting different ones. There are doubts about the validity of the tests themselves (do they measure needed skills or do they mainly measure skill in taking tests?). There is the fact that some skills are clearly more measurable than others, and that some most highly prized intellectual characteristics (creativity, ingenuity, motivation) are hard to measure at all.

5. Measures of Student Attitudes and Satisfactions

Many people feel strongly that intellectual skills are only part, perhaps not even the most important part, of the output of education. They believe that much of the value of education, to the individual as well as to society, lies in its effect on attitudes and personality. Educated people are more likely to read newspapers (which presumably contribute to their knowledge of public issues) and more likely to vote. There is some survey evidence that educated people are happier—more satisfied with

[11] For a discussion of why these results might occur, see Alice M. Rivlin, *Systematic Thinking for Social Action*, Washington, D.C., Brookings, 1971, pp. 70–78.

themselves, their marriages, and their jobs.[12] They are less likely to be criminals (or at least less likely to be caught!) and less likely to become public dependents.

A successful educational program presumably contributes to a student's general ability to function in society, although little progress has been made in measuring this ability. Some work has been done on self-image or self-worth measures on the hypothesis that a person with a positive view of himself and his capacity is more likely to function well.

In evaluating schools or teachers, remarkably little attention has been paid to direct measures of student satisfaction. One might think that the most obvious way to measure whether a school was performing well would be to ask the students how they liked it. More sophisticated attempts might be made to develop indexes of enthusiasm or motivation of students by observing their behavior. Except at the college level, where individual professors are sometimes rated by student questionnaire, however, little serious attention has been given to student satisfaction. Such a gap is surprising—especially in view of evidence that students learn more and perform better intellectually if they are happy—and may reflect some puritan moral judgment that education ought to be painful to be good.

6. Measures of Income Increase

From an economist's point of view, the most interesting effect of education is that it increases an individual's productive capacity—a fact reflected in the higher earnings of educated people compared with less educated people with the same personal characteristics. By now, there exists a voluminous literature on the theory of human capital and ways of measuring these income differentials. Vocational education and manpower training programs are frequently evaluated by comparing groups that went through the program with supposedly similar groups that did not, measuring the difference in expected future income, and comparing this discounted income differential with the cost of the program.

There are many difficulties with using income increase as a measure of output of education: (a) It is not easy to disentangle the effects of the education from the effects of other characteristics that influence income—such as ability and family connections—many of which are highly correlated with education. This is similar to the problem that arises with the use of test scores. (b) It is difficult to know how much of the income

[12] Stephen B. Withey et al., *A Degree and What Else? Correlates and Consequences of a College Education*, A Report Prepared for the Carnegie Commission on Higher Education, New York, McGraw-Hill, 1971, Chap. 5.

differential is attributable to actual skills and knowledge acquired by education and how much to the possession of diplomas, degrees, or other educational credentials. (c) Much education is bought for other reasons than to increase income, and it is difficult to separate the amounts or kinds of education whose objectives are primarily investment. Strictly "vocational" education may be plausibly classified as primarily designed to increase income, but only a small part of the education system admits to being primarily vocational. (d) Finally, there is the practical problem that the income increase attributable to education takes a long time to show up; income differentials computed from the experience of those who have passed through the education system many years ago provide no plausible clue to the system's current productivity, especially at the younger ages.

To improve the information base for personal and social decisions about education, it would clearly be useful to have a variety of measures of the effects of education, especially the effect on intellectual skills and knowledge, attitudes and satisfaction, and productivity or income. But the question here is what measure of performance, if any, should be built into the national accounts in order to obtain better measures of the changing productivity of the education sector comparable with those for other sectors of the economy. The answer, I am afraid, is that none of these approaches now offers a procedure which would be clearly superior to the current practice of taking the value of input (deflated by changes in input prices) as a measure of output.

If one used body counts, one would be denying the possibility of qualitative change in education. Historically, one would show a sustained decline in the productivity of American education as real resources per student, and especially teacher-student ratios, have risen. This decline may, of course, be a fact, but there is no reason to assume it. Attainment measures, as pointed out above, do not differ significantly from body counts.

Exposure measures, such as credit-hours or student hours, do appear to have some appeal, especially if the exposure unit can be weighted for quality (cost) as suggested above. Such a procedure might at least make the deflation of educational expenditures more comparable to the deflation of expenditures for other types of goods produced in the private market. It would not, however, get away from the fact that one is essentially using inputs to measure outputs.

The output measures which appear independent of inputs (test scores, attitudes, income differentials) all present difficulties. Besides the primi-

tive state of the art of measurement and doubts about the validity and reliability of individual measures, there is the overriding difficulty of weighting these measures. It is hard to believe that any set of weights that might be attached to sets of test scores, self-image measures, and income differentials would command enough consensus to be used as an output measure in national accounts. Even if it did, how would the results be used?

HOW-CAN-WE-DO-BETTER QUESTIONS

It seems useful to distinguish three rather different questions: (1) Is performance in education at least potentially measurable? (2) Would it be useful to develop better measures? (3) Would it be useful to build these measures into the present national income and product accounts, or some improved version of these accounts, in order to measure changes in education productivity? My own view is that the answer to the first two questions is a resounding "yes," and that the answer to the last question, at least for some time to come, is probably "no."

Only educational mystics persist in believing that the output of education is inherently and irrevocably unmeasurable. In practice, almost everyone (including teachers) behaves as though acquired skills and knowledge were eminently measurable. They give tests and impose education exposure requirements and base very important decisions on the results. They give more responsibility, better jobs, higher pay to people with more measurable "education" on the assumption that the performance of these individuals will justify this faith in measurement.

Of course, what is ordinarily being measured is a combination of the results of an individual's education with his inherent characteristics and other influences on his tested performance. However, there seems to be no inherent reason why these influences cannot be at least partially sorted out, especially if one is willing to make greater use of experiments than has been done in the past. Much of the difficulty in separating out the influence of education comes about because most of the data being used is retrospective and nonexperimental.

It is important to develop better performance measures in education. Indeed, it is hard to see how intelligent decisions about education can be made without them. Schools have been amazingly reluctant to provide any information on performance—no matter how it is measured—but taxpayers and parents are now demanding it. They want to know what they are buying and whether schools are getting better or worse. Test scores and other performance measures are now being used as evidence

against educators. It seems likely that educators will respond by developing more comprehensive and reliable measures of their own, not only to satisfy the public, but to put their own house in order and build into the management of education some measures of what is being produced and some incentives to produce it more effectively. Several current innovations in education may hasten this process. Performance contracting is based on the idea that output is measurable. Voucher systems and community control of the schools both seem likely to lead to a greater emphasis on performance measures. The consumer with a voucher will need school output measures so that he can comparison shop on a more intelligent basis. Communities that gain control of a local school will want some basis on which to evaluate the performance of the management.

It is easy to think of decisions that would be illuminated by having better performance measures for education available to the decision makers, even if several different measures were offered at once without weights. Decisionmakers or voters could supply their own weights; at least they would have a better notion of the choices. But it is hard to think of a decision that would be improved by building such performance measures into the national accounts, even if one could agree on the weights.

In our decentralized education system, no one actually makes decisions about the aggregate resources devoted to education; but even if someone did (say, if education were federalized) he would not find average productivity or even average rate of return estimates much use. Comparisons of performance of different schools or projects or programs would be useful in channeling resources toward the most effective ones. Estimates of rates of return to particular types of vocational or professional education programs might affect decisions. Information on the distribution of educational subsidies among groups in the population would illuminate equity or distribution decisions. All of these types of information would have higher priority for decisionmaking than building performance measures for the education sector as a whole into the national accounts.

COMMENT

BURTON A. WEISBROD, University of Wisconsin

The papers by Alice Rivlin and Mancur Olson cover a wide variety of issues associated with the process of rational decisionmaking in the public

sector in general and the education sector in particular. Olson directs his remarks toward two questions: why is there so much inefficiency in government, and what can be done about it? But there are two logically prior questions that must be dealt with. Are governments inefficient? And what, precisely, does "inefficient" mean?

Olson tells us, in effect, that since everyone knows that governments are inefficient, we can proceed to ask why. But surely, any serious assessment of "inefficiency" in government must begin with a definition of that term. The absence of a clear definition by Olson, and the plethora of cloudy statements, leave the reader quite unsure of what is meant by inefficiency. A sample of statements in the paper indicates that Olson includes in the term inefficiency such matters as "ineptitude or waste," and the absence of "Pareto-optimal efficiency." What does he mean when he asserts that "government production is generally less efficient *in some sense* than private production . . ." (emphasis added)? Does it make no difference in what sense, and compared to what, such inefficiency exists? Does Olson mean that governments are less efficient than (a) private sector firms "in general"; than (b) private sector firms that actually provide public-type goods; or than (c) private sector firms would be if they provided public-type goods?

Not only is the reader left uncertain as to what the "problem" is that Olson is addressing—since inefficiency is such a slippery term—but even if that were settled, another question would remain: Are governments inefficient? My own judgment is that we have little evidence to which we can refer. Olson's references to the "apparent" failures of government to reach the production frontier, or to the view that "schools are widely thought to be inefficient," and to the judgment of "anyone who has traveled in the Soviet Union as well as the West . . . [that] many of the services provided by the Soviet government seem to be relatively inefficient" are rather weak evidence.

The primary goal of his paper, as I see it, is an exploratory analysis not of the reasons for government inefficiency, but of some important differences between the goods and services governments produce and those that nongovernmental units produce. Olson regards the key difference, and hence the key problem, as "the lack of measured output" of the government sector. This, in turn, is the consequence of the government's "provision of collective goods and the handling of externalities."

Actually the problem of output assessment in the public sector is, I believe, decomposable into three subproblems: (1) defining targets or goals—that is, deciding which effects of a particular program are cared

enough about to be regarded as "outputs"; (2) developing satisfactory operational measures of outputs—that is, of movement toward those targets—measures that do not affect output by the process of measurement; and (3) valuing each output.

One point to underscore, and Olson states it clearly, is that many of the problems he notes would not disappear—although they might or might not diminish—if the government turned over to the private sector the responsibility for producing and distributing many of the goods and services now in governmental hands. I suggest that this is particularly relevant to the choice between governmental production of a good and governmental provision of it via contracting-out, with or without subsidization. Since a government can ordinarily achieve its objectives of allocative efficiency or distributional equity through either its own production or through stimulation of nongovernment production, it is not clear why and under what circumstances governments opt for one or the other. There has been little attention paid to this matter in the literature; more is needed.

The first subproblem is defining goals, or What is an output? Consider a case in which consumers prefer different forms of each collective-consumption good but they agree that, given the costs of production, or the desire for equity, only one form or "quality" will be provided governmentally. When this is the case, each consumer will attempt to obtain the type or form of public service he prefers, for consumers will disagree as to what is a "relevant" output. Differences in preferences can be interpreted as differences in objectives and, hence, as differences in what each consumer regards as an output. Thus, the political struggle over the form of public service—e.g., in education, health, transit, and so forth —reflects disagreement over what is an "output," where an "output" from the viewpoint of any consumer is something which enters positively into *his* utility function. In short, some public service may produce output for you but not for me, or vice versa. Since identification of outputs thus depends on utility functions, the difficulty of defining outputs should not be surprising.

Developing operational measures of output—i.e., of progress toward goals—is the second problem associated with output valuation in the public sector. Assume that agreement has been reached on what we mean by "outputs" of any given government service, and that operational measures are being developed. A key problem at this stage is that the particular measure used to monitor (or measure) output is likely to influence the behavior of production units in unintended ways. For

example, if the "output" of a school (the subject of Alice Rivlin's paper) is measured by performance of children on a standardized test, teachers will have an incentive to "teach to the test." In a recent experiment with "performance contracting" in U.S. public schools—an arrangement by which private firms are paid by government according to improvement in students' reading and mathematics achievement—an attempt was made to circumvent this problem by choosing randomly from among five or six different tests of achievement in each subject. There is, however, a dilemma: either the results of all the tests in a particular subject are highly correlated, in which case a teacher who teaches to any one test teaches to them all, or the results are not highly correlated, in which case the measured output of a teacher's or a school's efforts will depend on which of a variety of equally suitable measures were used. Of course, if we had great faith in all the tests, then we would not care if the teacher taught "to" them.

Since Olson emphasizes the importance of developing operational measures of all government outputs, it is interesting to examine further the kinds of operational measures for education that were surveyed by Alice Rivlin. Lying behind the choice of measures is the issue to which I referred above: What do we want from the government service? In other words, what do we regard as an "output"?

In general, we economists tend to think of outputs in value-added terms. In the case of education, however, most of the proposed measures of output that Alice Rivlin enumerates reflect no attempt whatsoever to distinguish between the level of a student's accomplishment and the addition to that level that results from schooling, ceteris paribus. Grades on tests and in courses, for example, measure, at best, what students know, not what the addition to their knowledge has been, and in no way do they reflect the separate influence of schooling variables as distinct from ability, motivation, etc. This suggests that tests do not generally measure "learning added" by schooling.

But by contrast, what if the objective and, hence, the "output" of schooling were in terms of certifying competence of students, as an aid to matching workers and jobs in the labor market? In that case, information about level of achievement would truly represent a value added, regardless of whether any of the achievement was attributed to schooling. It appears that the education system may be measuring its outputs in ways that imply that certification is a most significant output, while what is learned in, and because of, school is much less significant. My point is not to take a position on which output is more important—learning or

certification—but to note the relationship between the measure used and the implied notion of what is an output. I agree with Alice Rivlin that an index of changes in educational output over time is needed, but I would add that the usefulness of such an index will depend critically on how satisfied we are that it measures the outputs we care about—i.e., outputs that help us to achieve our social goals.

As the problems of output measurement in the public sector are considered, it is useful to note what our measures of outputs in the private sector do and do not measure. Consider, for example, Alice Rivlin's statement that in the case of education, "some highly prized intellectual characteristics (creativity, ingenuity, motivation) are hard to measure at all." I suggest that we distinguish between the problems of measurement and of forecasting. "Creativity" and "ingenuity" of a person are extraordinarily difficult to predict, but are not nearly so difficult to measure as of a given point in time. The same is true, I believe, of private sector outputs. Durable goods prices reflect current assessments of expected outputs; how accurate the buyer's predictions turn out to be is a matter about which we seldom inquire. Need we do so in the public sector? And if we do, should it be said that the public sector is inefficient compared with the private unless we know how well the private sector predicts and how well it would predict the effects of activities in which it is not currently engaged (but has left for the government)?

Valuation of outputs is the third problem associated with output assessment in the public sector. Olson believes that the problem of inefficiency in the public sector is largely attributable to the collective-consumption nature of government services. It is surely a considerable oversimplification, however, to suggest that governmental provision of goods and services is largely limited to collective goods. Most governmental goods are rather far removed, in fact, from the pure collective-good case. And perhaps more importantly, much of what governments provide, and the manner in which such goods and services are provided, relate to distributional objectives. Zero-pricing, for example, often reflects, I believe, not the view that marginal social cost of an additional consumer-user is zero, but rather that a zero price provides equality of access to the good or service. Some recent research has raised questions about the fact of equality of access, but this does not vitiate the argument that many people believe that setting prices at zero enhances "equality of opportunity." In any case, zero-pricing of goods for which marginal social cost of a consumer is positive—perhaps because of congestion costs—is bound to pose allocative efficiency problems. Thus, it

seems that governments may be trading off allocative efficiency for distributional equity. Whether the trade-offs being made are or are not efficient (in the Pareto sense) is a question worthy of further study.

The frequent use by governments of zero-pricing poses one problem, and it points up another problem in connection with our efforts to value government outputs. First, the absence of observed market prices makes it difficult to estimate output values; and second, governmental efforts to bring about greater equality of access than would occur if market demand considerations prevailed points up the distributional objective of government activities. The presence of this latter objective suggests that market-clearing prices, even if we could know them, would not be the "right" prices to use! It is for this reason that my sympathy for Mancur Olson's interesting proposal to run experiments to discover consumer demand functions is somewhat tempered.

If the relevance of distributional goals for evaluating government performance is granted, we might wish to conclude that the "same" outputs differently distributed are, in effect, different outputs or, at least, are different "values" of outputs. Then, since public and nonpublic provision of a service is likely to be distributed differently, it may not be so obviously correct that, as Alice Rivlin put it, "it clearly makes no sense to treat the costs of public and private educational institutions differently in the national accounts." Perhaps it is not entirely senseless to regard the same resources, when directed toward different consumer groups, as having different values.

One conclusion does, however, seem clear: many hurdles remain to be surmounted in the process of defining the relevant outputs of the government sector, developing operational measures of the outputs, and then valuing increments of those outputs.

Zvi Griliches, Harvard University

Rivlin's suggestion that different outputs in education should be weighted by relative costs is interesting and a step forward. To implement it, however, we would have to separate, among other things, the cost of university training into its undergraduate, graduate, and research components. But that would not be easy, as anyone who has thought about it a little knows. Moreover, we would have to worry about the comparability of particular completion levels over time. There are conflicting trends here that should be looked into. Structuring our estimates

this way would force the "quality" question into the open and might result in our getting some additional information on it.

Olson claims that public enterprises are inefficient because they are in the business of producing output which is very hard to measure. I am not sure that this is either necessary or sufficient for the inefficiency charge. My feeling is that the problem of inefficiency is contained in the statement that they are not maximizing the right thing, that the internal incentive structure is wrong. Whether this is a consequence of the difficulty of measuring the "output" that we would want them to produce is not all that obvious. There are many private enterprises that produce hard-to-measure output in this sense but do not seem to be subject to the same inefficiency problems.

The difficulties of measurement are real, and they apply with special force to the problem of measuring the contribution of investment in research to economic growth. They can be illustrated by some recent work of my own.[1] To estimate the contribution of such expenditures to the growth in total factor productivity we need (besides the observable level of these expenditures) three numbers: (1) The fraction of such expenditures that could conceivably have an impact on productivity as currently measured; (2) the fraction of such expenditures that can be thought of as increasing the stock of relevant knowledge (i.e., what fraction of this investment is net?); and (3) the social rate of return to such investments. A careful look at the sources and uses of these funds convinces me that no more than about a half of these expenditures could affect measured total factor productivity in the United States. A large fraction of these expenditures is spent on space exploration and defense, where the achievements, if any, are valued at cost. At the same time, scattered evidence on the longevity of research results (the bulk of which is in the applied area) suggests that a depreciation rate of about 10 per cent per annum may not be out of line, particularly at a time of relatively high levels of gross investment. That would imply roughly that about half of the observed expenditures could be treated as net investment. There remains then the question of the relevant rate of return. Since we are interested in the social rate of return to such expenditures, we cannot estimate it from observed profit and loss statements. We are thus forced either to an explicit calculation of social rates of return for selected innovations or into econometric analyses of interindustry differences in productivity growth. A survey of earlier results on the

[1] Z. Griliches, "Research Expenditures and Growth Accounting," International Economic Association Conference paper, St. Anton, Austria, 1971.

former and some experimentation with the latter leads me to the use of 30 per cent as an upper bound on the average rate of return to all the relevant research and development expenditures. Since the over-all ratio of R&D expenditures to private domestic GNP was 0.03 in 1970, these numbers taken together imply a contribution of about 0.2 per cent per year to the rate of growth of total factor productivity ($0.3 \times 0.5 \times 0.5 \times 0.03 \simeq 0.002$). This is not negligible but neither is it overwhelming.

While each of these numbers is based on some evidence and quite a bit of scrutiny on my part, I hope that I have conveyed implicitly the wide uncertainty surrounding each one of them. While they are about as good as I can make them at the moment, I would not urge the OBE to take over their routine production and to award them their explicit Good National Accounting Seal. I think such exercises are interesting and worthwhile but are based on too sparse a data base and too weak a theoretical framework to warrant their incorporation into the official accounts as they are currently constituted.

NESTOR E. TERLECKYJ, National Planning Association

I would like to discuss the measurement of output which would reflect the use and usefulness of goods and services *after* they have entered the household sector. Such measurements are important for an understanding of a number of current issues in that they may make it possible to judge the "cost-effectiveness" of consumer expenditures in terms of the capacity of consumer goods (and public programs) to satisfy specific wants—e.g., for shelter, health, or mobility—as represented by specific criteria or indicators in a manner similar to the methods employed in analyzing the effectiveness of public expenditures.

Measurement at this level of utility is much more difficult than the measurement of the public sector output alone as discussed by Schultze, but the nature of problems encountered is similar in both cases. He pointed out that in the private sector valuations of commodities are given by the market, but that in the public sector comparable decision relevant valuations are not available and that the measurement of physical units is not likely to reflect the really important performance characteristics. These are more likely to be the functional characteristics of goods and services measured by their capacity to contribute to the output of a "consumer production" process.

But, in attempting to evaluate the prevailing effectiveness of the satisfaction of consumer wants, one necessarily must consider a much wider set of data than that generated by the operation of the markets because both market and nonmarket goods enter as inputs into their satisfaction and there is no reason to expect that optimizations occur between the market and the nonmarket components. For this reason there is much less of a basis than in some other contexts to rely on the valuations produced by the existing markets as necessarily applicable to the more fundamental and in a large measure nonmarket concerns.

In dealing with what Olson calls "things that people really care about," the really important indicators of performance may be quite different from the usually measured physical units or the factor cost and price, mainly because of the nonlinearities in the production of consumer satisfactions and the nonexistence of the equimarginal optimization in the large (over the given system of inputs).

It is important to attempt to define output of the final demand sectors in terms better related to welfare concepts than the amount of resources used in its production. The units of measurement of performance characteristics that people care about may be defined by specific indicators for the consumer units to reflect actual objectives of consumption. It may be an interesting question for future research to find out whether and to what extent those performance elements can be aggregated into some kind of social indicators representing national totals. Perhaps the work that Hartle described can suggest some possibilities.

The joint relationships among various inputs and outputs, both public and private, can perhaps best be seen as an incremental input-output matrix, with the input dimensions representing specific goods, services, and public programs and the output dimensions representing indicators of satisfaction of specific wants and with interaction effects among the inputs. For example, the output of an automobile in terms of mobility and access without the public complements of roads and highways is negligible by itself, and in many cases may approach zero. Its value, in a consumer decision sense, also depends on the extent and quality of the existing and prospective (over the lifetime of the commitment made) alternative systems.

Thus, the "performance output" of a given consumer or public program is contingent on the level of the complementary inputs. The "success" of a highway will depend on what type of industrial or residential development will occur; the extent to which a given public health program may be successful will depend on how the public and the medical

professions will respond to it; an educational innovation at the elementary level may critically depend for its success on the response of the parents, etc.

In the sense of utility generating consumer activities then, any government program, like any private product, is only a partial input, and almost any nontrivial level of performance output and any efficient mode of its production requires complementary "inputs" from both sectors. This suggests that the magnitude of the "performance product" (total product) of a single governmental program per se is inherently indeterminant, because of the jointness in production of the higher order output within the final demand. All we can really determine ex ante is the range in which the output of a public program can fall, when measured from the given initial conditions and as contingent on the complementary inputs, or ex post what has been the joint result of a combination of events, in particular periods and circumstances, and what range of magnitude of effects of particular factors is suggested by the evidence.

Finally, performance measurement has to be selective and relativistic. It cannot and probably should not be definitive. To the extent that a given public program or a given market good or a combination of market goods and public programs has multiple outputs, the conflict about defining *the* correct measure of output can be reduced, if not resolved, by measuring more than one important performance characteristic where there is more than one. A matrix representation is quite helpful in this respect. This would also reduce the risk of bias in policy making, which would automatically follow from selecting a single performance characteristic for attention.