EDUCATION AND PRODUCTION FUNCTIONS

TOWARDS AN EDUCATIONAL PRODUCTION FUNCTION •

SAMUEL BOWLES • HARVARD UNIVERSITY

1 INTRODUCTION

EDUCATION occupies an important position in every major economy of the world. In the United States over 6 per cent of gross national product is annually spent on formal schooling alone, and the amount is increasing at a rate more than twice that of the economy as a whole.[1] According to Machlup's estimates for the year 1958, the resource costs of education and training, broadly defined, amounted to over 12 per cent of the value of GNP.[2] Education is called upon to accelerate the rate of growth and to equalize the distribution of income. In the poor countries schools are regarded as a central element in the economic infrastructure. In the United States, schooling and training programs receive the lion's share of the funds of the war on poverty. Everyone seems to have accepted James Mill's dictum that "if education cannot do everything, there is hardly anything it cannot do." The growing popular interest in education has been paralleled by the development of an immense literature on the role of human capital in economic growth and the distribution of income. And yet nobody really knows how education is produced.

NOTE: I should like to acknowledge the competent research assistance of Matthew Lambrinides, and the helpful comments of Susan Contratto, Herbert Gintis, Christopher Jencks, Henry Levin, Arthur MacEwan, Stephan Michaelson, Christopher Sims, Lester Thurow, and Thomas Weiskopff. In preparing this paper, I received financial support from the U. S. Office of Education and the National Science Foundation.

[1] Council of Economic Advisers [15], p. 143. The figures do not include foregone earnings; they refer to the year 1966 and the previous decade.

[2] Fritz Machlup [49], pp. 354 and 362. The resource costs of education and training include foregone earnings, and expenditures on training on the job, as well as in the army and other institutions. The gross national product figure is adjusted to include items counted as educational costs, but not normally included in the national income accounts.

This paper is a discussion of the educational production process. An educational production function is the relationship between school and student inputs and a measure of school output. This method of representing the educational production process is likely to be of particular value both in descriptive studies of human capital formation and in normative investigations to determine optimal educational resource allocation.

If schooling does have a unique effect on labor productivity or earnings, we should be able to trace this effect to the development of cognitive skills and attitudes in school. We may also be able to relate the development of productive skills and attitudes to school policies concerning the allocation of scarce resources. A production function relating school inputs to the development of productive capacity should give us a better indication of why the more educated are better qualified for productive roles. Moreover, differences in production functions for different racial and social class groups, as well as differences in educational inputs for these groups, may help explain an important aspect of the determination of the distribution of personal earnings.

In setting school policy and in long-range educational planning, knowledge of the educational production function is essential to efficient resource allocation. This is true whether the decision unit is pursuing the objective of growth or of equality, or a combination of these and other (perhaps noneconomic) goals. Without an estimate of the technology of education (the production function) the relationship between the opportunity cost and expected benefits of particular policies must be little more than guesswork.

In this paper, I will concentrate on the relationship between school inputs and conventional measures of school outputs, such as achievement scores. An explanation of the relationships between scholastic inputs and achievement, on the one hand and economic performance on the other, would have been more germane to the theme of this conference—education and income. In fact, much, though not all, of the economic relevance of what follows depends on there being some systematic relationship between scholastic achievement and economic performance. For the present, however, I can only present results based on the educational evaluation of the schools' outputs.

An educational production function is defined as follows:

$$A = f(X_1, \ldots, X_m, X_n, \ldots, X_v, X_w, \ldots, X_z) \tag{1}$$

where

A = some measure of school output—for example, a score on a scholastic achievement battery;

$X_1, \ldots, X_m$ = variables measuring the school environment. The variables here would typically include the amount and quality of teaching services, the physical facilities of the school, and the length of time that the student is exposed to these inputs;

$X_n, \ldots, X_v$ = variables representing environmental influences on learning outside the school—e.g., the parents' educational attainment; and

$X_w, \ldots, X_z$ = variables representing the student's ability and the initial level of learning attained by the student prior to entry into the type of schooling in question.

We are interested in estimating the structural parameters of the function f. It will be seen later that, although we cannot estimate the equation in the form presented, some progress can be made with a slightly modified version.

The data at our disposal are ordinarily based on a cross-section of students. Although I dwell at length on the deficiencies of our particular data, the information generally available for the purposes of estimating educational production functions is, in some respects, superior to that available for estimating economic production functions: we have data at the "firm" level and therefore avoid the problem of making "technological" inferences based on industrial, state, or national averages; most of our input data are measured directly, rather than in monetary aggregates; and we have ample data on the quality of the factors of production—e.g., teachers, principals, and other school personnel.

The crucial deficiency is not in the lack of data but the absence of a theory of learning to guide us in establishing a model for our estimation. The engineer can tell us exactly what technical processes are necessary for the production of particular physical commodities; these processes, in turn, suggest appropriate specifications of the production function, as well as some a priori limits on plausible estimates. In education, the psychologist replaces the engineer or agronomist as the source of technical information on the production process, and, despite fruitful developments in learning theory, we still know very little about the underlying technology. Nonetheless, a reasonable a priori model of the production of scholastic achievement *can* be specified on the basis of existing theory, and preliminary estimates based on this model are encouraging.

Although attempts to measure the relationship between school inputs and outputs have occupied the attention of a number of educational researchers over the last half-century, the estimation of the structural parameters of a production function similar to (1) is a relatively new approach.[3] The results of these studies are difficult to summarize, in part because of the variety of measurements used, and because of the diversity of findings. In any case, the purpose of this paper is not primarily to present empirical estimates of production functions but rather to explore some of the conceptual and econometric problems involved in this type of estimation. The results of some of these studies, as well as my own results, are included as examples.

Section 2 includes a discussion of the behavioral assumptions underlying the usual production function estimates and the particular difficul-

[3] Herbert Kiesling [45, 46] used data generated by the Quality Measurement Project of New York State to estimate school production functions for various communities in New York. Martin Katzman [43] estimated production functions for a variety of school outputs of elementary schools in Boston. As a part of the study which gave rise to the report of the Central Advisory Council for Education (the Plowden Report) in England, G. F. Peaker estimated a series of production functions for British elementary education. Thomas Fox, John Holland, and Jesse Burkhead have estimated production functions for a wide range of school outputs for Atlanta and Chicago, as reported in Burkhead [9]. I have not included in this list the study of Finis Welch [66], as he relies on highly aggregated inputs and his estimates can only be identified as educational production functions by some stretch of the imagination. Eric Hanushek [30] and David Armor [2] have used U.S. data on the sixth grade to estimate production functions for elementary education. The International Project for the Evaluation of Educational Achievement, under the direction of Torsten Husén, has estimated similar functions for the determination of mathematics achievement in a sample of twelve countries [36]. A considerable amount of additional work is now in progress. Larry Posner (at Harvard) is currently estimating production functions for on-the-job training.

ties encountered when such concepts are applied to schools. Sections 3 and 4 are devoted to the measurement of school outputs and student inputs. The measurement and interpretation of school inputs are discussed in Section 5, and some preliminary illustrative results are presented in Section 6.

2 ESTIMATING A PRODUCTION MODEL FOR SCHOOLS

IN A statistical investigation using nonexperimental data, the most we can expect is to discover some relationship among measurable dimensions of the process, based on the particular configuration of the data in our sample. In this we are limited both by the preconceptions of the researchers who selected the sample and obtained the data and by the patterns of variation which school decision-making processes have brought about in the sample of schools chosen. To use the apt analogy of Marshak and Andrews [50], we are in the position of neither the agronomist nor the meteorologist. The agronomist seeks to understand production relations in agriculture with a mind to increasing productivity; he can experiment, varying his inputs systematically and in any desired combination and thus, under ideal conditions, predict the effect on productivity of specific changes in inputs. The meteorologist relies on nonexperimental data, but seeks only to predict normal behavior—not to affect events. We have the worst of these worlds. We seek to affect educational output by altering school inputs, and our data are generated by decision-making units and student responses entirely beyond our control. Thus our ability to calculate the consequences of changes in existing educational processes is very limited indeed.[4] But the limited ability to vary and control school inputs which occur in our data is just one of many difficulties.

[4] A considerable amount of educational research has used experimental techniques. See, for example, Gray and Klaus [25] and Kirk [47]. These methods hold some promise for empirical determination of the educational production function, particularly when we seek to estimate the consequences of major departures from existing technologies.

Assume for the moment that we seek to estimate the production function (1) in the form:

$$A_i = f_0 + f_1X_{1i} + f_2X_{2i} + \ldots + f_zX_{zi} + u_i \tag{2}$$

where

A_i = the achievement score (or other output measure) for the i^{th} student;

$f_0, \ldots, f_z$ = the parameters of the production function to be estimated;

X_{ji} = the amount of input j devoted to observation i's education, $j = 1 \ldots z$; and

u_i = the disturbance term.

The least squares technique yields unbiased estimates of the regression coefficients only if the independent variables, X_{ji}, are exogenous—i.e., only if they are uncorrelated with the disturbance term, u_i. We may, however, expect the school inputs to be endogenous to some system—for example, a system of equations based on the school administrators' preference function, the educational production function(s) and an educational budget constraint. In this case, we are faced with the problem of simultaneous equation bias which plagues the estimation of production functions at the firm level;[5] a single equation approach to the estimation of (2) will yield inconsistent estimates of the structural parameters f_j.

The basic implausibility of the above behavioral model provides one way out of this difficulty. Given that school administrators know very little about the underlying production processes and are subject to a wide variety of political and legal constraints, we can assume that they do not select or alter school inputs with a mind to optimizing any well-defined function of school outputs. Therefore, we can assume for purposes of estimation that the X_j's are exogenous.

Rejection of an optimizing decision model for school administrators alleviates at least one simultaneity problem (there will be others), but it deprives us of the usual interpretation of the estimated parameters of (2) as a production function—a relationship which in conventional usage indicates the *maximum* output consistent with a given set of inputs.

[5] See Marshak and Andrews [50] and Nerlove [54] for a discussion.

If school administrators follow no systematic optimizing behavior, then the observations on which our estimates are based are not generally technically efficient. (It may be possible to produce the same output with less than the observed level of all inputs.) Thus we arrive at some sort of *average* production function. Only if the absolute degree of technical inefficiency is uncorrelated with the level of factor inputs (which seems unlikely) will the estimates of f_j from (2) represent unbiased estimates of the true production relationship.[6]

Moreover, the school output is multidimensional, and the relative valuation of different outputs—say, mathematical competence as opposed to citizenship—differs among school districts. For this reason, technical inefficiency may result neither from inadvertence nor from the absence of optimizing behavior but rather from the conscious pursuit of objectives not adequately measured in any single index of school output. The appearance of technical inefficiency in this context can occur whenever school administrators are able to vary the composition of the school output by exercising choice about the allocation of *a given set* of resources within the school. As long as the levels of factor inputs measured in our production function do not completely determine the composition of school output, the failure to produce the maximum possible level of one dimension of output with a given set of inputs may reflect a relatively low valuation of that output by the school personnel.

A perfectly analogous problem arises because students differ in the aspects of the school output which they most successfully acquire. Not all students seek to maximize their scholastic achievement. For this reason, we may observe low achieving schools whose technical inefficiency results from the fact that the students have chosen to emphasize the noncognitive aspects of personal development. Given the possibility of making this kind of tradeoff, it is not surprising that McDill, Meyers, and Rigsby [52] found that a measure of the degree of intellectualism and achievement orientation in the predominant value systems of schools is a good predictor of scholastic achievement.[7]

[6] Of course, the constant term will be biased downward. If we had a number of different observations on inputs for the same school, perhaps from different grades, or years, or tracks, we might be able to use school dummy variables to eliminate this "management bias." See Massell [51] and Hoch [32].

[7] Of course, it is virtually impossible to establish the causal relationship between intellectual values and scholastic success.

Differences in student valuation of scholastic success and other dimensions of output present serious difficulties in the measurement of output as well as the interpretation of the estimated production function. We use the score on any achievement test as an output index to measure some underlying competence purportedly learned in school. However, in order for the test to be an adequate measure of the level of competence, it must be the case that children try equally hard to do well on the test, or more simply that they tell all that they know. Yet to some children the reward of doing well on the test is by itself sufficient to call forth the maximum effort, while for others the testing situation does not elicit full concentration. Zigler and DeLabry [72], for example, found that, although there were significant differences in middle-class and lower-class students on a set of tests given in normal conditions, when each group was tested under the reward conditions which had been found to be optimal for that group, there were no group differences in performance. Likewise, Terrell, Durkin and Wiesley [61] found that material-reward conditions produced better performance in lower-class children, and nonmaterial reward proved more effective with middle-class children. Thus the importance of the social class of the student's family in our production functions is probably not a pure measure of a direct influence on learning, but represents a proxy for motivation in test-taking as well. If this is the case, the structural parameters of (1) relating to the student's social background will be upward-biased estimates of the effect of social class on school learning.

While some school inputs can perhaps be regarded as exogenous to our system, one set of inputs—student attitudes—must be endogenous. Student attitudes toward self and toward learning are a consequence of past and present achievement (as well as other influences) and are important determinants of achievement. Thus we have to rewrite equation 2 as

$$A = f(X_1, \ldots, X_n, \text{attitudes}) \tag{3}$$

and,

$$\text{Attitudes} = g(X_1, \ldots, X_n \text{ achievement—past and present}) \tag{4}$$

In this case simultaneity seems unavoidable. Estimates based on (3) will certainly be biased, since attitudes will in general be correlated with the disturbance term. The solution is to return to (2). Although this equation does not directly incorporate attitudes (the explanatory variables being those which are exogenous), it incorporates the effects of attitudes indirectly, as they are related to the set of exogenous variables. Unless we are specifically interested in increasing scholastic achievement by changing student attitudes directly, we lose little if we exclude attitude variables from the equation.[8]

The dearth of knowledge concerning the learning process makes any a priori specification of form for the estimation of educational production relationships particularly difficult. The notion of diminishing marginal product is appealing, although not well established in the field of education. A function linear in the logarithms of the variables would seem somewhat superior, particularly in view of the possibility of positive interactions between inputs. Nevertheless, the restrictions of the Cobb-Douglas function are severe.[9] An analysis of variance designed to identify the nature of the interaction between inputs would seem a prerequisite for the adequate specification of the form of the educational production function. In the work below, I will use the simple linear additive form presented in (2) above.[10] I have not been able to compare my results with those generated with alternative forms of (2).

Children do not learn in the same way, nor do they learn the same things. Lesser and Stodolsky, for example, found dramatic differences in the patterns of scholastic proficiency on four different dimensions of learning among Chinese, Jews, Negroes, and Puerto Ricans.[11] When we find consistent differences in patterns of response to school inputs, we have good grounds for grouping students according to these patterns and

[8] Nonetheless, in my results I present both the reduced form and the (biased) estimates of the structural equation including attitudes.

[9] Particularly important, to my mind, is the fact that the cross derivatives among any pair of inputs, each of which is positively related to output, must also be positive. This would require, for example, that an increase in the quality of teachers be more effective on children of well educated parents than on the children of illiterate parents.

[10] The choice is dictated largely by data considerations. My regressions are estimated from correlation tables, not from the raw data. Thus I was unable to make the necessary logarithmic transformations. Hanushek [30] found that the logarithmic form gave slightly better significance for the estimates of the parameters of his production functions.

[11] Stodolsky and Lesser [60].

estimating a number of different technologies. Casual inspection of the results of Hanushek's, Kiesling's and my own work suggests that it is useful to think of distinct educational production technologies—at least for the four-way classification of students—black/white and rich/poor.[12] The need to stratify the population by race and class reflects the primitive level of our inquiry. In a completely specified model the differences in the behavior among these subpopulations ought to be attributable to characteristics of these groups which are relevant from the standpoint of learning theory.

Based on recent findings in the study of economic growth, we may anticipate that the major changes in productivity of schools will be effected not by a more rational input structure within the existing technology, but by changes in production functions themselves—including changes in the relationship between home background and achievement, as well as the more conventional input-output relationships. If our goal is to effect such changes, we should identify "best practice" schools and attempt a quantitative explanation of their superiority.

3 THE MEASUREMENT OF OUTPUT

WHAT DO schools produce? They perform two primary economic functions: selection and socialization. The socialization process may be broadly construed as the preparation of youths to fill adult roles. This involves the transmission of skills and perhaps even more important, the indoctrination of values and commitments appropriate to successful adult participation in life. Many dimensions of school output are directly relevant to economic performance, others are valued for different reasons. Both are economically important.

We would like to measure school output by post-school economic or social performance or by indexes of valued characteristics thought to be acquired in school. Unfortunately, our indexes of school output are based largely on tests administered in school and designed to measure

[12] However, I have seen no statistical test of the hypothesis that the underlying subpopulations differ significantly. I have estimated functions for black twelfth-grade students separately, as well as black twelfth-grade students stratified by region.

scholastic achievement. These achievement scores must, then, be considered proxies for, or, perhaps, influences on, later economic behavior. Scholastic achievement is, presumably, not valued per se, but only as input to subsequent measures of performance. Therefore, although we use achievement, A, as the measurement of output, our rationale for this is that many socially or individually valued characteristics are themselves functions of scholastic achievement. As an illustration, we may consider the capacity to earn income, E, as such a valued attribute and investigate the function:

$$E = E(A). \tag{5}$$

There are indications (at least for some groups) of a significant relationship between scholastic achievement and earnings. Wolfle [71], for example, found that among persons with the same number of years of schooling, percentile class rank in high school, and other measures highly correlated with school learning (IQ scores) showed positive relationships to annual earnings. Weisbrod and Karpoff [67] found a strong relationship among college grades and earnings in a large sample of employees of a nationwide firm. When Weiss [68] measured years of schooling in achievement units rather than calendar years, the significance of his estimates of the relationship between earnings and schooling improved.[13] Hansen, Weisbrod, and Scanlon [29] also found a highly significant relationship between earnings and scores on the Armed Forces Qualification Test—a measure highly correlated with school learning; and Duncan [19] reports similar findings for a somewhat more representative sample.[14] The apparent unimportance of scholastic achievement in the

[13] His study refers to white male workers in the North Central region of the United States, and is based on the 1/1000 sample of the 1960 Census. Years of schooling for each subcategory of workers were translated into achievement years based on evidence in Coleman [13], on mean "years" of achievement for groups of individuals classified by place of schooling (urban, rural, north, south, etc.). An "achievement year" is a norm based on the average achievement scores attained in each year of school by white students in the urban northeastern United States.

[14] Some additional evidence is surveyed in Griliches [26]. It should be pointed out that in the Hansen-Weisbrod-Scanlon [29] sample of draft *rejects*, the relationship was remarkably small; and in Weiss' study of black workers, only *one* age group exhibited a significant relationship between years of schooling (measured in achievement units) and earnings. Weiss reported similar results using a direct measure of years of schooling completed.

determination of earnings of some categories of workers (black, for example) may indicate that our present schools are economically ineffective for these groups. However, it is also possible that other dimensions of school outputs such as aspects of personality development not reflected in achievement scores have a more direct bearing on economic performance. In this case we have selected the wrong dependent variable in our investigation.

Most educational processes may be thought of as producing intermediate goods for use in other educational or training activities. Virtually all workers receive some on-the-job training, above and beyond the general education received in school. And most forms of schooling are a direct input into yet higher levels of education. In the empirical work presented below, we are particularly interested in education at the twelfth grade. At the present time in the United States, about half of the high school graduates move on to further education. Thus the evaluation of the school output must not be confined to the direct effects of schooling on the productivity of the worker—the effects on the efficacy of vocational training and further schooling must also be considered. Here we have strong evidence that success in vocational training as well as in higher education is significantly related to the initial level of scholastic achievement.[15] It may be that the main economic importance of scholastic achievement in secondary school is not its direct contribution to production, but rather its effect in increasing the "trainability" of workers and consequently in reducing the costs of further human capital accumulation at the post-secondary school level.[16]

Scholastic achievement, of course, is not the only dimension of school output. Literally hundreds of tests have been devised to measure "achievement" alone, and this is only one aspect of the effect of schooling on cognitive skills and personality. In addition to economic performance in the post-school years, schooling may affect an individual's self-concept and his sense of control over his environment. These and other aspects of personality development may be valued per se, and additionally may be important determinants of post-school economic performance.

[15] Jensen [39], Ghiselli [24] and Astin [3].

[16] In the above discussion of these issues, I have drawn heavily on the unpublished work of Herbert Gintis.

TABLE 1

Zero-Order Correlations Among Measures of School Outputs, Twelfth-Grade Boys, United States

	2	3	4	5	6
1. Information total	.23	.65	.76	.54	.19
2. Self-confidence		.17	.19	.09	.11
3. English total			.67	.46	.26
4. Mathematics total				.57	.20
5. Abstract reasoning					.19
6. Clerical checking					

NOTE: Based on a sample of 3,027.
SOURCE: J. C. Flanagan, *et al.*, *Project Talent,* Pittsburgh (1962), Table 2–7j. (Description of test scores accompanies table.)

It is safe to say that there are fewer independent dimensions of school output than there are test instruments to measure them. But if we rely on the (unsatisfactory) evidence of zero-order correlations among individual test scores (see Table 1), we find that the relationship between some of these measures is rather weak.

Thus the output of schools is multidimensional with a vengeance, and to complicate matters, there are no convenient sets of "prices" with which to aggregate the output. Moreover, the technologies for the production of each dimension of the output are blatantly dissimilar. For example, my estimates of the reduced-form equation (2) in which the dependent variable is scholastic achievement, differ considerably from estimates in which the dependent variable is an index of the student's sense of control over his environment.[17] My preliminary results indicate that for twelfth-grade black students in the United States the verbal ability of teachers is the most important explanatory variable in the former equation, while the racial composition of the school and the experience of the teaching staff is more important in the latter.[18]

[17] The measurement of this variable is described in the appendix to this paper.

[18] In the latter equation teacher experience and the proportion of school enrollment which is black are positively related to a crude measure of the student's sense of control over his environment.

Apparently, then, schools are multiproduct firms; and the composition of output is highly sensitive to the particular combination of inputs used. The school production function must be represented by a number of equations, each relating the school inputs to a different dimension of output. The choice of an optimal input structure thus depends on the relative valuation of the different school outputs and on the rates of transformation among these outputs implicit in this system of production equations.

For the purposes of educational policy making, we are particularly interested in the structural parameters of production function (2), for under ideal conditions they may be interpreted as the marginal products of the inputs in question—that is, $MP_j = \partial A/\partial X_j = f_j$. If we know the social opportunity costs of inputs, p_j, we may use this information to move in the direction of optimal input proportions as defined by the conditions[19]

$$\frac{\partial A/\partial X_j}{\partial A/\partial X_k} = \frac{\hat{f}_j}{\hat{f}_k} = \frac{p_j}{p_k} \qquad \text{(for all pairs, } j, k\text{)} \tag{6}$$

However, difficulties arise when we seek to compare the marginal products of the same input for two different groups of students. We find, for example, that the estimate of the structural parameter relating to the verbal ability of teachers as an input into an achievement production function is considerably greater for black twelfth graders in the United States than for whites. Can we infer from this that verbally adroit teachers ought to be shifted from white to black districts? The answer is no.

The output measure is ordinal; there is no zero point and no well-defined unit of measurement for achievement.[20] Thus, while the marginal rate of substitution in production—represented in the additive linear form by the ratio of regression coefficients of any two input factors—is still a valid analytical concept, the absolute magnitude of the marginal product is not. Among students scoring at very different parts of the

[19] Of course, we are here accounting for only one output.

[20] At least one writer has constructed a cardinal index of achievement based on the size of vocabulary Bloom [5], pp. 103–04. Whether words known is linearly related to anything important is not known.

scale of measurement, equal units of increase in scores are not comparable; for example, it may be "easier" to make gains at the lower end of the scale than at the upper end due to a so-called "ceiling effect."[21] We really need to know the relationship between our output measure A and measurements of directly desired performance, such as earnings. The studies by Weiss, and Hansen, Scanlon and Weisbrod, mentioned earlier, suggest a linear relationship between their measurements of achievement (in one case achievement years, in the other the AFQT score).[22] Although this evidence is encouraging, it is certainly not sufficient to justify much confidence in a cardinal interpretation of academic measurement of school learning.

A further problem remains. Our output indexes are subject to some error—that is, test score = "true measure" + error, and, consequently, var(test score) = var(true measure) + var(error). We have no idea of the validity of the test—that is, its correlation with a hypothetical true measure. But some idea of the magnitude of the error may be gained from estimates of the reliability of the tests.[23] The reliability of our tests is in the neighborhood of .9. Taking this as an upper estimate of the validity, at least 19 per cent of the variance of the test scores is due to test errors. Assuming that the errors in test measurement are uncorrelated with our explanatory variables, even if our explanatory variables predict the true measure with perfect accuracy, a validity of .9 imposes an absolute maximum proportion of variance explained by our equations of .81. It will be seen below that the actual R^2's are considerably lower.

[21] "Most frequently, aptitude and achievement tests are constructed in such a way that it is harder to secure significant changes on one part of the scale than on another. This unevenness stems from the combined effect of a ceiling on the tests as well as the greater difficulty of the test items which can make the difference at the high end of the scale." *Ibid.*

Also, Chausow [11] indicates that among a relatively homogenous group of individuals, students who are initially high on a test characteristically make smaller gains than students who are initially low.

[22] For white males Weiss found that the linear relationship was superior to a logarithmic or polynomial one.

[23] Although there are various ways of measuring test reliability, we may convey the essential meaning as the zero order correlation between scores on the odd and even number questions of the same test or the zero order correlation between two versions of the test given to the same individual at roughly the same time.

4 THE VALUE-ADDED PROBLEM

AN ACHIEVEMENT SCORE must be considered a measure of *gross* output. Our goal is to estimate the relationship between school inputs and *net* output, or value added. For this we need a measure of the raw material inputs, i.e., student ability, or, alternatively, the level of learning upon entry to the school in question. (Without such a measure our efforts are like attempting to measure the effectiveness of a beauty parlor without knowing what the clientele looked like to begin with.)

The problem is that all measures of relevant student "ability" are hardly distinguishable from measures designed explicitly to test scholastic achievement.[24] According to a survey of the evidence by Bloom, simple correlation coefficients between intelligence and achievement scores (when both tests are administered at the same age) are ordinarily in the neighborhood of .85.[25]

There are a number of possible interpretations of this close association between measured achievement and intelligence. First, it may be that the tests simply measure the same thing. There is strong evidence that intelligence as measured by the usual instruments is a developmental concept, measuring general learning.[26] Moreover, most IQ tests depend heavily on verbal facility, which is probably a good reflection of general school learning.[27] A second explanation is that the tests measure different dimensions of competence, but that they both are sensitive to variations in the school environment. Evidence that abilities as measured by IQ tests are significantly influenced by the educational environment is available in the data from a study of identical twins separated prior to the age of three and reared apart (see Table 2). Results not reported there showed that over 60 per cent of the variance in the IQ differences between paired identical twins can be explained by differences in the educational environment; alone, the differences in the physical and social environments

[24] Duncanson [20] and Cohen [12].

[25] Kelley [44], pp. 193–213; Bloom [5], pp. 102–03; Coleman and Cureton [14]; Duncanson [20], and Cohen [12]. All of these studies use correlations corrected for test reliability. Jensen [39] reports lower correlations, although no references are given.

[26] See Hunt [35].

[27] Bloom [5], pp. 71 and 104.

TABLE 2

The Effect of Environmental Differences on IQ Differences Among Paired Identical Twins Reared Apart

Environmental Difference	Effect[a]	t Statistic
Educational	.66	4.2
Social	.25	1.6
Physical	.19	1.3
R^2:	.70	
$\lvert X' X \rvert$:	18	
Number of observations:	18	

[a]Normalized regression coefficient of the environmental difference in an equation predicting I Q differences.

SOURCE: Data from F. N. Freeman; H. H. Newman; and K. J. Holzinger; *Twins: A Study of Heredity and Environment,* Chicago, 1937.

explain less than a third of the variance.[28] A third interpretation holds that there is little casual relationship between intelligence and achievement, but that for some reason the bright children go to good schools, receive especially sympathetic attention of school personnel within their schools, and consequently achieve well. Although evidence on this interpretation is fragmentary, data from the U.S. Office of Education's Equality of Educational Opportunity Survey suggests that skepticism is in order. The relationship at grade one between a measure of verbal ability and the two school inputs found to be most important in the determination of scholastic achievement is very weak. The zero order correlations between verbal ability and a measure of the teacher's verbal ability are .05 and .02 for blacks and whites, respectively; the correlation between student verbal ability and a proxy for the adequacy of the school's physi-

[28] The evaluation of the social, physical, and educational environments was a subjective assessment by a panel of judges who were not cognizant of each others' evaluations or of the twins' test scores. Although the judges were in close agreement, we have little knowledge of what they took into account in their evaluation. See Freeman, *et al.* [23].

cal plant was −.07 and −.02. Moreover, in an equation including measures of the parents' level of schooling and other social class dimensions no school input measures were significantly related to grade-one verbal ability among black students.[29] Finally, it is possible that achievement and intelligence tests measure distinct abilities but that intelligence is the primary influence on the acquisition of scholastic knowledge.

Of course, elements of each interpretation are consistent with parts of the others. What is important here is that, to the extent that either of the first two explanations is correct, it is illegitimate to include a measure of IQ in the production function as an independent variable, as the importance of the school input variables will thus be underestimated. Note, however, that if *both* the third and fourth interpretations are correct, the exclusion of the intelligence measure from our equations will bias upwards the estimated effect of school resources.

On balance, the evidence seems strong enough to reject the use of a contemporaneous IQ score as the measure of the student "raw material" input into the production process. What are the alternatives? As we are interested in measuring school learning, it would seem reasonable to use tests of learning administered at grade one as a measure of raw input. Because these first grade tests clearly measure the combined effects of genetic ability and environmental influences prior to age six, they are exactly what we need. Thus, our basic equation (1) becomes

$$A_{12} = f(X_1, \ldots, X_v, A_1) \tag{7}$$

where subscripts on the achievement variable refer to the grade at which the test is taken. In order to estimate a function of this type, we need individual test scores for students at two different levels of schooling. While some data of this type are currently available, and more is on the way, we are generally forced to rely on cross-sections.

If (7) is the correctly specified relation, and we are forced to work with data which do not include the first grade scores (A_1), we may be able to estimate the unbiased regression coefficients of (7) if we have independent evidence on $b_{1.12}$, the regression coefficient of A_1 in equation 7, as well as the estimated equations:

[29] The data are from volume II of Coleman [13].

$$A_{12} = f^{12}(X_1, \ldots, X_v) \tag{8}$$
$$A_1 = f^1(X_1, \ldots, X_v) \tag{9}$$

The unbiased estimates of the regression coefficients of (7).
are then

$$b^*_j = \hat{b}_j^{12} - b_{1,12}\hat{b}_i^1 \tag{10}$$

where $\hat{b}_j^{12}$, $\hat{b}_j^1$ are the estimated regression coefficients of X_j in equations (8) and (9), respectively. This approach is equivalent to Theil's method of estimating the bias due to specification error.[30]

I have assumed that the relationship between first grade and twelfth grade scores is such that a student scoring one standard deviation above the mean at grade one will, *ceteris paribus*, score .5 standard deviations above the mean at grade twelve. Thus,

$$b_{1,12} = .5\left(\frac{\sigma_{12}}{\sigma_1}\right) \tag{11}$$

where σ_1 and σ_{12} are the standard deviation of achievement scores at grades one and twelve, respectively. This figure is somewhat arbitrary. It is based on two sets of data. First, longitudinal studies of scholastic achievement scores suggest a simple correlation between early and late test scores in the neighborhood of .6 to .9. Most of the studies cover substantially less than twelve years, so we may suspect that the simple correlation of scores at grades one and twelve would be somewhat lower.[31] Moreover, the simple correlation is not the appropriate evidence, as we seek an estimate of the partial effects of differences in A_1 on A_{12}. To the extent that students who initially score high on tests are exposed to a better learning environment in either home or school, the size of the

[30] Theil [62]. Our method is based on the assumption that the function f accurately represents the relationship between each X_j and first grade scores which prevailed at the time of entry into school, and that the vector $V_1, \ldots, X_r$ is the same for a given student at grades one and twelve.

[31] Based on forty-one longitudinal achievement score correlations reported in Bloom [5], pp. 106–09. The correlations for more widely separated years occupy the lower end of this range.

above-reported correlations exaggerates the normalized partial relationship between initial endowments and later scholastic achievement.[32]

A study by John Conlisk, using longitudinal data on students' intelligence scores, provides additional evidence.[33] Using a sample of seventy individuals in Berkeley, California, Conlisk estimated the following regression equations:

$$\underset{}{IQ_{18}} = \underset{(6.44)}{4.77} + \underset{(.099)}{.490}\ IQ_{1-5} + \underset{(.358)}{1.514}\ \text{YrsSch} \qquad R^2 = .45$$

$$\underset{}{IQ_{18}} = \underset{(5.74)}{8.11} + \underset{(.093)}{.527}\ IQ_{6-8} + \underset{(.367)}{1.051}\ \text{YrsSch} \qquad R^2 = .49$$

where $IQ_{t\text{-}u}$ = score on IQ test administered between ages of t and u, and YrsSch = number of years of school attended.

The IQ measures are standardized indexes with identical means and standard deviations; the standard errors of the estimated regression coefficients are in parentheses. Although the correct equation for our purpose would predict scholastic achievement and would include measures of school inputs, the biases are likely to be small unless there is a strong association between the early IQ measure and the quantity of school resources.

We assume that the function f^1 (equation 9) will consist entirely of arguments relating to the social class and home background of the student, since school inputs could hardly affect scores on tests taken at the beginning of grade one.[34] For this reason, correction of this specification bias will involve reductions in the estimates of the coefficients of variables which measure the student's social class and home environment. Some difficulties in implementing this correction for specification bias will be discussed in the concluding section.

32 All of the achievement measures are subject to error. At grade 1 the reliability of the achievement score used (verbal ability) is .78. If the validity of this score is only slightly below its reliability, the portion of variance in A_1 due to random error is .5. Thus our method is equivalent to assuming that the normalized partial relationship between the *true* measure of initial endowments and A_{12} is 1.

33 Conlisk kindly allowed me to use his unpublished results.

34 There is ample evidence that grade one achievement scores are associated with measures of student social class. See Bereiter [4] and Gray and Klaus [25].

5 THE INPUT STRUCTURE OF THE SCHOOL AND OTHER LEARNING ENVIRONMENTS

WE WANT to estimate the effect of school inputs on the value added of schools. In order to isolate this effect, however, we must specify, as fully as possible, all the environmental influences on learning—home, community, peer groups, and school. A complete specification of the model is particularly important in view of the specification bias likely to arise because of the close statistical association usually found between school and home environments that are highly conducive to learning.

We may derive some suggestion of the relative effects of various dimensions of environment on learning from another study of identical twins reared apart. In this case the differences in paired identical twins' scores on Stanford Achievement Tests are the measure of differential learning; the relationship between environment and learning based on this study of twins is illustrated in Table 3. Comparing the contents of Table 3 with Table 2, we see that the educational environment is of

TABLE 3

The Effect of Environmental Differences on Scholastic Achievement Differences Among Paired Identical Twins Reared Apart

Environmental Difference	Effect[a]	t Statistic
Educational	.899	7.69
Social	.024	0.21
Physical	.001	0.01
R^2:	.82	
$\|X'X\|$:	.86	
Number of observations:	19	

[a]Normalized regression coefficient of the environmental difference measure in an equation predicting achievement differences among paired identical twins.

SOURCE: Based on data of Freeman, Newman, and Holzinger (see Table 2).

considerably greater importance in the explanation of achievement differences than of IQ differences. Educational environment alone explains more than 80 per cent of the difference in scholastic achievement between paired identical twins. While this is hardly surprising, the insignificance of the social and physical environment among genetically equivalent individuals is striking.[35] Of course, this may well be due to the imperfect measures of the environments in question, but it alerts us again to the dangers of specification bias in equations which do not include some measurement of initial endowment and suggests that much of the importance of social class to school learning apparent from cross-section studies may reflect genetic differences associated with the educational and social characteristics of the student's family.

Let us begin by asking which aspects of the student's environment could have some effect on learning. A brief survey of the literature on learning suggests that the major environmental influences on school achievement (in addition to general intelligence) include:

a. The quantity of verbal interaction with adults[36]

b. The quality of verbal interaction with adults[37]

c. The motivation for achievement and understanding in the environment[38]

d. The richness of the physical environment.[39]

The available measures of these dimensions of the environment are far from adequate. However, data exist which allow us to attempt an empirical implementation based on the above a priori specifications. Moreover, a number of relations in sociological and psychological research will assist us in implementing the model.[40]

[35] Alone they explain only .13 of the variance of achievement differences.

[36] Anastasi [1].

[37] For example, see Olim, Hess, and Shipman [55], and Jackson, Hess, and Shipman [37].

[38] See Dave [16].

[39] Particularly for very poor children, the level of family income may have a strong causal relation to the development of intelligence and scholastic achievement. See, for example, Harrell, *et al.* [31].

[40] There are some grounds for believing that the student's peers exercise an effect on learning. I have not included this discussion of the learning environment as I have been unable to find adequate measures of the peer group environment in the data which I am currently using.

Let us begin with the nonschool environment. We may use a measure of parental education to represent the quality of the verbal interaction between child and adult;[41] and family size (as well as the number of adults living at home) provides a rough measure of the quantity of interaction and communication.[42] If we restrict ourselves to variables which can be regarded as largely exogenous, the motivation for achievement may be measured in terms of parental attitudes toward schooling[43] as well as in terms of the potential objective importance of education to the student. Race, for example, may constitute a logical measure of expected returns: we have compelling evidence that the economic returns to schooling, at the elementary and secondary levels, are significantly less for black than for white children.[44] The physical environment of the home may be measured by the quantity of reading material in the home, the parents' occupation or income, or proxies for these variables, such as quantity of consumer durables, etc. Evidence of a relationship between malnutrition (primarily protein deficiency) and learning difficulties suggests that some measurement of the physical environment may substitute for a measure of the physical development of the child as related to learning, particularly for very poor children.

A number of authors have attempted to account for the familial and social environment by stratifying their analyses according to social class.[45] Available evidence suggests that while this technique is certainly useful in reducing the multicollinearity among the explanatory variables, it is a thoroughly inadequate representation of nonschool effects on learning. Peterson and DeBord [57], for example, found that *within* two refined

[41] On the importance of language models, see Olim, Hess, and Shipman [37], and Jackson, Hess, and Shipman [55].

[42] Anastasi [1].

[43] Although we are not able to include this variable in our analysis below, as we have no adequate measures in our sample, at least one study which sampled the parents as well as the children, has confirmed the importance of parental attitudes. See Peaker [56]. Of course, parental attitudes toward schooling must depend in some degree on the particular school in which the child is enrolled. Thus parental attitudes are not unambiguously exogenous.

[44] See Michaelson [53], Weiss [68], Hanoch [28]; some of the data are summarized in Bowles [6]. Differences in family interest in schooling and its associated impact on children's motivation is, in part, a cultural phenomenon, likely to vary among ethnic groups. For convincing evidence in one case, see Gross [27].

[45] For example, Kiesling [46] and the U.S. Commission on Civil Rights [63], in their study of the effect of racial integration on scholastic achievement.

TABLE 4

A Model for the Estimation of the Environmental Influences on Learning

Underlying Influence on Learning	Empirical Representation in the Model: Home	Empirical Representation in the Model: School
1. Quality of verbal interaction with adults	a. education of parents	a. educational level of teachers b. other measures of teacher "quality", such as verbal-ability c. school policies d. teacher attitudes
2. Quantity of verbal interaction with adults	a. family size b. one or both parents absent	a. class size
3. Motivation for achievement in school	a. parental attitudes toward education b. race, ethnic group c. objective returns to schooling	a. community support of education
4. Richness of the physical environment	a. family income or occupation, consumer durables in the home b. reading material in the home	a. school facilities, labs, libraries, texts, etc.

substrata (white and black lower-class urban children in the South), variables measuring home environment and parent-child interaction explained .56 (white) and .66 (black) per cent of the variance in achievement scores.[46] The predictive power of dimensions of home environment *within* narrowly defined social strata suggests that an analysis using no other control for social environment will be subject to serious specification bias.[47]

We may proceed in roughly the same manner (although with less confidence) with the empirical measurement of the school environment. The quality of interaction between child and adult may be represented by a measure of the educational level or verbal proficiency of the teachers. The quality of the interaction may depend somewhat on school policy, which may be represented by a host of imperfect measures of such aspects of school environment as breadth of curriculum, amount of extracurricular activities, etc. (Although there is some evidence—e.g., Chausow [11]—that methods of teaching make a difference, the available data do not allow even a rough measure of this variable.[48]) The physical environment of the school may be represented by measures of special facilities—labs and libraries—as well as overcrowding, makeshift classrooms, and so on.

Table 4 summarizes our model of the environmental influences on learning.

Notice that even this partial specification of the learning environment includes fourteen measures, many of which are highly correlated. Thus serious multicollinearity problems arise. In the estimation of a full model of the type specified for twelfth-grade black students in the United States, the determinant of the $X'X$ matrix, which is probably the best

[46] Of course, the Peterson and DeBord findings could result from collinearity between the home environment and school inputs to which the children were exposed. This probably will not explain the entire result, however. Within a group of black sixth-grade students in the third socio-economic quartile, Levin found that in addition to various school input measures a number of home measures were significantly related to scholastic achievement.

[47] The strength of the measured relationship between school inputs and achievement observed by Kiesling may be due in part to this bias.

[48] For an exception, see Peaker [56], who found in many cases that an assessment of the teacher's proficiency by an inspector was significantly related to scholastic achievement.

single measure of the presence of multicollinearity, fell to .0005.[49] Similar problems arise for whites and for other grades. In order to estimate the above model, we need to reduce the number of variables so as to simplify the presentation and bring the multicollinearity problem within tolerable limits. That is, we would like to replace the equation

$$A = f(X_1, \ldots, X_v) \tag{12}$$

with

$$A = F[g_1(X_1, \ldots, X_v), g_2(X_1, \ldots, X_v), \ldots, g_h(X_1, \ldots, X_v)] \tag{13}$$

where $h < v$.

Thus we may wish to define a new variable, e.g., "teacher quality," as an aggregate of individual variables measuring verbal ability, years of schooling, experience, certification, and so on. If a significant degree of multicollinearity arises from intercorrelations *within* each set of variables forming an aggregate variable, the problem will be reduced; the new aggregate variables, represented by $g_1, \ldots, g_h$, may be sufficiently orthogonal to allow successful estimation of the relationship. The choice of a precise grouping of factors is determined by more than the desire to reduce multicollinearity; however, the usual aggregation rules do not seem particularly helpful here, as we have no knowledge of the matrix of second derivatives and cross-derivatives which would allow us to make use of them.

We have no previous results or compelling theory to use as a guide as to how to aggregate. In situations where all inputs are priced in the market, and where the assumption of maximizing behavior is somewhat more plausible, we ordinarily use factor or commodity prices as the basis of aggregation, as in the measurement of "capital" or intermediate inputs. Failure to appreciate the importance of these assumptions in the validity of any monetary aggregate in production theory has led to the frequent use of what might be called spurious factors of production, such as expenditure per pupil and teachers' salaries. In my own empirical work

[49] See Farrar and Glauber [21]. X is the matrix of normalized observations. The determinant of the $X'X$ matrix varies between 1, indicating complete orthogonality of the variables, and 0, indicating linear dependence of at least two vectors of observations. See also Bowles and Levin [7].

(for black twelfth graders), whereas teachers' salaries explain only .0085 of the variance in achievement, the two variables most closely related to variations in teachers' salaries—teachers' verbal ability and years of schooling—explain over four times as much.[50]

Similarly, Kiesling [46] found a "disappointingly weak" relationship between expenditure per pupil and achievement; in a later paper [45] he found that variables which together explain most of the variance in expenditure per pupil are very strongly related to achievement. All of this simply suggests that school administrators are not using their resources efficiently as far as the production of scholastic achievement is concerned. The relative market prices used to aggregate inputs in the expenditure measure are apparently significantly different from the marginal products of the inputs.[51] Thus the use of monetary aggregates is unfounded in theory.

In our situation the best method seems to be to attempt to identify the underlying dimensions of the input structure by both a priori and empirical methods. This done, we would select a variable, or an index based on a number of variables, to represent each dimension. Our a priori specification of the school environment suggests that we have four important dimensions of the school environment: teacher quality, teacher quantity, school policy, and physical facilities. One procedure would be to assume that these represent the dimensions of the input structure and to select from each set a variable to represent the underlying input. Thus we could represent teacher quality by the teacher's score on a verbal ability test (at least when we are predicting verbal achievement), and so on.

Alternatively, we may combine our preconceptions based on previ-

[50] In each case I am referring to the increase in the coefficient of determination in an equation already including measures of social background and nonteacher school inputs, as in equation 8. See Levin [48] for an analysis of the relation between teacher quality and teacher salary. These two teacher attributes (verbal ability and years of schooling) explain 60 per cent of the variance in teachers' salaries in the sample reported in the next section.

[51] This inference is supported by a comparison of the estimated marginal products f_i and the supply prices for various teacher attributes. (See Levin [48].) Calculations of the cost of unit increase in achievement through increases in each factor based on these estimates show that for the sample under consideration increases in teacher's verbal ability are more efficient than any other dimension of teacher quality, by a wide margin.

ous research and learning theory with an empirical analysis of the structure of our data, using principal components analysis.[52]

Leaving the problem of aggregation in this unsatisfactory state, let me ask how well we have measured the environmental influences on learning, particularly as they relate to the school. The answer "not very well" stems primarily from three problems: (a) the home and school variables fail to capture the complexity and richness of the interaction processes relevant to learning; (b) we have ignored significant qualitative differences in education available *within* a school; and (c) we have measured inputs at only one point in time, while the learning process must certainly be cumulative and therefore dependent on past inputs.

Turning to the first problem, our measures of social class, family size, class size, teacher quality, and school facilities do not measure the quantity and quality of interaction as relevant to learning. Our input measures are merely circumstantial evidence of a few of the opportunities for such interaction. Two recent studies suggest that these crude measures are a poor substitute for measures of actual patterns of interaction. On the basis of detailed interviews with sixty parents, Dave [16] and Wolf [70] found that their measurement of home environment explained .57 and .64 of the variance in intelligence and achievement, respectively.[53] The crude home environment measures used in our study explain only 10 per cent of the variance in individual achievement scores. Presumably, analogous detailed studies of actual classroom interaction would reveal that our school measures are an equally poor representation of our basic learning model.

The second problem arises particularly where tracking is widespread and the differences in the educational opportunities within a single institution are so great that we really have two or three schools in one.[54] Also, differences in teacher and administrator attitudes and expectations may differ considerably within a school and even within a classroom.[55]

[52] This is the method used by Kiesling.

[53] Recall, also, the Peterson and DeBord (1963) study, *op. cit.*

[54] Differences in the quality and quantity of school inputs received within the same school are documented in Hollingshead [34].

[55] See Davis and Dollard [17], pp. 284–85, and Warner, Havinghurst and Loeb [65], as well as more recent studies by Deutsch [18] and Wilson [69].

One recent study (Rosenthal and Jacobson [58]) suggests that teacher expectations have a significant effect on learning, at least in the early years of school. Failure to measure these in-school and in-class differences in inputs results in a specification error which is particularly serious because of the correlation of these differences with other of our explanatory variables. Because low social class and minority racial or ethnic status are closely associated with intraschool deprivation of school inputs,[56] the estimates of the parameters reflecting the impact of social class and race are biased upward. Further, because of the serious errors introduced by the school-wide aggregation of the variables measuring school inputs, the estimated effect of the school environment is biased downward.[57]

Our third objection, against the sole use of contemporary input measures, would not be serious if children did not move from school to school and inputs were roughly uniform throughout all of the grades up to the one for which the production function is being estimated. Of course, the world is simply not like that, and I think we sometimes underestimate the seriousness of this problem. In a sample of black sixth-grade students in a Northeastern metropolis, 57 per cent had attended more than one school since grade one, and 29 per cent had attended more than two.[58] Evidence from a number of studies of the phasing of learning development over the school years suggests that this problem is particularly serious, as patterns of achievement are apparently established with a high degree of stability in the early grades. Scannell [59], for example, found that scores on fourth-grade tests (Iowa Tests of Basic Skills) explained half the variance in test scores (Iowa Tests of Educational Development) in the twelfth grade.[59] Cardinal measures of scholastic achievement based on vocabulary tests suggest that about two-thirds

[56] See the evidence in Hollingshead [34] and the more recent studies cited in Rosenthal and Jacobson [58].

[57] In a study in which within-school variations were measured, Peaker [56] found that school inputs were considerably more important in the determination of school achievement (relative to other influences such as home background) when within-school variations in these inputs were taken into account.

[58] Work in progress by Henry Levin and Stephen Michelson.

[59] Scannell [59]; Bloom [5] summarizes the evidence on the stability of achievement.

of what is known in grade twelve was already known in grade six. On the presumption (which seems to have currency among educational psychologists) that the effects of environment on learning are potentially greater during periods in which the most learning takes place, it would seem that measurement of the inputs in the early grades would be essential to the prediction of achievement at the higher levels.[60]

The importance of the early years in the learning process suggests one last question: how much impact can we expect schools to have on learning?

The mental ability commonly called intelligence is probably the single most important determinant of scholastic achievement. Studies of the degree of hereditability of the characteristics measured by scholastic achievement tests are, to some extent, contradictory—the very wide range of estimates does not allow any simple summary of the results.[61] Nonetheless, it is safe to conclude that a sizable portion of the variance in scholastic achievement is associated with genetic differences. Thus, even in an otherwise perfectly specified and perfectly measured model, unless we are able to take account of genetic differences, we are likely to be unable to account for all of the variance of scholastic achievement. Of course, if the genetic component in intelligence is related to social class (through inheritance), our social background measures in the learning model will take account of some of the genetic differences in mental abilities among our students. Thus, the residuals in the estimation of the education production function cannot be unambiguously identified as the influence of student ability differences.[62]

Given the importance of genetic influences on scholastic achievement the possible impact of schools on achievement is severely limited. But

[60] In the absence of a time series of school inputs, it might be advisable to concentrate on the estimation of production relations in the early grades.

[61] See Jensen [39], Burt [10], Vandenberg [64], and Jensen [38].

[62] There is evidence of a significant genetic component in the observed social class differences in measured intelligence. For example, the IQ's of children adopted in early infancy show a much lower correlation with the social class status of their adopting parents than do the IQ's of children reared by their own parents; the IQ's of children reared in an orphanage from infancy and who have not known their parents show only slightly lower correlations with their true father's occupational status than that found for children reared by their own parents. See Jensen [40], pp. 1–2 and references.

this reasoning still overstates the importance of schooling, as the classroom is only a small part of the learning environment of the child. During the elementary and high-school years, children ordinarily spend considerably less than one-fourth of their waking hours in school. In addition, Bloom [5] suggests, about one-third of adult learning is achieved before age six.

Jensen [39] suggests that the impact of well-designed and well-executed "enrichment" and "cognitive stimulation" programs is ordinarily between 0.5 and 2.0 standard deviations on specific achievement tests. These rough estimates place broad limits on the expected maximum effect of either a very good or a very bad school as opposed to an average one. Of course, our estimations are based on a cross-section of representative schools, not well-endowed experimental programs designed to raise scholastic achievement. This fact plus the severe deficiencies in the measurement of the learning environment suggest that the impact of extreme school environments estimated from our production functions are likely to fall short of Jensen's estimates.

6 AN EDUCATIONAL PRODUCTION FUNCTION

THE FOLLOWING results are presented here to illustrate the discussion in the previous section of educational production functions. Their empirical substance must be treated with extreme caution.

The estimates are for black students enrolled in the twelfth grade in the fall of 1965. The data were collected by the U. S. Office of Education as part of the Equal Educational Opportunity Survey. Some of the results of this survey have been reported in *Equality of Educational Opportunity,* known popularly as the Coleman Report, after its principal author.[63] The sample and a number of serious shortcomings of the data are described in detail elsewhere.[64] Any reader adventuresome

[63] Coleman, *et al.* [13]. Our estimations are based on the correlation tables and mean and standard deviation of each variable, as reported in Vol. II of *Equality of Educational Opportunity.*

[64] In addition to the report itself, see also Bowles and Levin [7], Hanushek [30], and Hanushek and Kain [42].

enough to take seriously my preliminary results is urged to consult these sources.

The variables used in the empirical implementation of the model, along with their means and standard deviations, and a table of zero order correlations appear in the appendix. A listing of all other variables which were used in the experimental stage also appears there. A total of thirty-six variables were tested for significance in the educational production function. The equations presented below include all of the variables which

TABLE 5

An Educational Production Function (Reduced Form), Black Twelfth-Grade Students

Independent Variable[a]		Regression Coefficient (t in parentheses)	Beta
1. Reading material in the home		1.9284 (2.5847)	0.0822
2. Number of siblings (positive = few)		1.8512 (4.3411)	0.1316
3. Parents' educational level		2.4653 (4.4660)	0.1431
4. Family stability		0.8264 (1.6938)	0.0494
5. Teacher's verbal-ability score		1.2547 (7.1970)	0.2222
6. Science lab facilities		0.0505 (2.5821)	0.0784
Constant:	19.4576 (5.1887)		
R^2 :	0.1708		
$\|X' X\|$:	0.6628		
Number of observations:	1,000		

[a]Dependent variable is verbal achievement.

made sense on the basis of the learning model, and which proved to be significantly related to achievement.

The estimate of our basic equation of the educational production function (8) appears in Table 5. Note that the estimated parameters are consistent with our suggested model of learning. With one exception, all of the estimates, including those for the school environment, are significantly different from zero at the 99 per cent level.[65] The quality and quantity of interaction with adults, as well as motivation for schooling and the richness of the home environment, are all represented (at least symbolically) in the four nonschool environment variables.

Note also that the estimated parameters of the school inputs are consistent with the suggested model of learning. The very significant estimate of the influence of teacher quality (represented by the teacher's verbal ability score) is not surprising, since the teacher is the single most important school input.[66] The importance of teacher quality, which our estimate demonstrates, has also been confirmed by other work on educational production functions.[67] The failure of the class-size variable to appear in the equation may reflect severe errors in the measurement of this variable.[68] Although a number of studies have suggested that class size is not a significant influence on achievement (e.g., Hanushek [30] and Levin [work in progress]), Kiesling did find a highly significant rela-

[65] The family-stability variable is significant only at the 90 per cent level. The regression coefficients of the reduced-form equation, excluding attitudes, is virtually unaffected by the removal of the absent father variable.

I have fewer schools than individuals, and because of the strong possibility of there being unmeasured school effects, the variance-covariance matrix of the error term is not diagonal: the off-diagonal elements reflect the covariance of the residuals among students in the same school. As a result I have probably underestimated the standard errors of the estimates of the regression coefficients. Without school identifications for each observation, I am unable to estimate the extent of this bias.

[66] The teacher's verbal ability test consists of only 30 questions and is self-administered. If, as seems likely, the variance of the error component in this measure is larger than the variance of the error component in the dependent variable, the estimate of the associated regression coefficient may be biased seriously downward. The same reasoning, of course, applies to the other inputs.

[67] See Kiesling [45]; Hanushek [30]. In addition to these results, Levin found that two measures of teacher quality (verbal score and type of college attended) were highly significant in explaining verbal achievement among sixth-grade black students of the third socio-economic quartile in a large Northeastern metropolitan area.

[68] See Bowles and Levin [7].

TABLE 6

Educational Production Function (Reduced Form), with School Policy and Community Support Proxies, Black Male Twelfth-Grade Students

Independent Variable[a]	Regression Coefficient (t in parentheses)	Beta
1. Reading material in the home	1.6579 (2.2193)	0.0707
2. Number of siblings (positive = few)	1.7583 (4.1322)	0.1250
3. Parents' educational level	2.4519 (4.4575)	0.1423
4. Family stability	0.8339 (1.7174)	0.0499
5. Teacher's verbal-ability score	1.0419 (5.5605)	0.1845
6. Science lab facilities	0.0373 (1.8824)	0.0580
7. Average time spent in guidance	1.4803 (2.3652)	0.0804
8. Days in session	0.2032 (1.9213)	0.0582

Constant:	-14.2214 (-0.7529)
R^2 :	0.1804
$\|X'X\|$:	0.4477
Number of observations:	1,000

[a]Dependent variable is verbal achievement.

tionship between students-per-teacher and achievement.[69] It is somewhat more surprising that the very crude representation of the physical facilities of the school—science laboratories—appears to be significantly related to achievement.

Note that the explained variance is very small. This is to be expected, given the crudeness of our measures, and it points to our failure to specify adequately a model of school achievement. In addition to the poor measurement of our variables we have certainly omitted altogether some important influences on learning.

The absence of a measure of school policy, which would help to indicate the quality of student-teacher interaction, is explained by the profusion of imperfect measurements of this input dimension. When we entered eleven school policy and environment variables (see Table A.1) into the above equation, we could not accept the hypothesis that the entire set of regression coefficients for these variables were zero.[70] In order to represent the influence of school policy variables, we have introduced a proxy variable, representing the extent of guidance counseling in the school. We have further added a days-in-session variable to represent the general level of community interest in and support of education. The resulting equation appears in Table 6.

Both of these proxy variables are highly correlated with measures indicating over-all support for education—e.g., teachers' salaries and system-wide expenditure per pupil. In addition, both are positively asso-

[69] My preliminary results with a different sample of black twelfth-grade students in the Northeast and Central United States reveal a significant negative relationship between class size and achievement. The positive relationship between class size and a measure of school output found by Welch [66] is almost certainly a reflection of the smaller classes in rural schools and the failure to take account of the negative influences on learning associated with a rural home and community environment. The negative association between student-teacher ratio and tenth-grade verbal scores in twenty-two Atlanta public schools estimated by J. W. Holland and J. Burkhead [9] is difficult to interpret, as the equation in which this finding is reported includes a measure of per pupil expenditure (plus a number of insignificant variables). This seems to suggest that even with a given level of expenditure, reduction in class size produces sufficiently strong effects on achievement to more than offset the associated opportunity costs.

[70] The F value leading to the rejection of the hypothesis was 2.39 with 11 and 984 degrees of freedom. Thus the hypothesis was rejected at the 99 per cent level of significance.

ciated with such school policy variables as extracurricular activities and foreign language courses (though days in session is less closely associated than is guidance counseling.)[71]

Regressions similar to that reported in Table 6 were estimated separately for different samples of 1,000 students each in the North and in the South. The estimated equations (which are presented in an appendix) are remarkably similar to those estimated for the national sample.[72]

Thus far we have been working with a model which takes no explicit account of students' endowments at the beginning of school. The resulting biases in our estimates are suggested by the following exercise, based on equations (9), (10), and (11). We have attempted to explain a similar achievement score in grade one using our set of explanatory variables. The resulting equation and the calculation of the specification bias appear in Table 7. At the first-grade level, coefficients of the school input variables were never significantly different from zero (at conventional levels).

Given the crudeness of both the measurements and the technique, the particular numerical estimates are subject to considerable error. The effect of correcting for this specification bias is to reduce the apparent influence of social class on school learning. Of course, as long as we use an additive linear model with no interaction effects and plausibly find no relationship between school inputs and initial scores, there can be no estimated bias of the school inputs.

It is likely that some of the remaining influence of social class and home background is a reflection of genetic differences. This is certainly a plausible interpretation of the results, given the apparent unimportance

[71] Of course, the days-in-session measure may simply reflect urban-rural differences, as there is evidence that rural schools are open for fewer days per year. (See Coleman [13].) As a test of this hypothesis, we added a variable measuring the size of the senior class to the equation. This new variable was insignificant, and, although its introduction lowered the estimated regression coefficient for days in session by about 10 per cent, the latter variable was still significantly different from zero at the 95 per cent significance level. The remainder of the equation was altered only slightly. The importance of guidance counseling is equally difficult to interpret, as an abundance of counselors may be associated with a large fraction of college-bound students in the school, or severe discipline problems, or both.

[72] In both North and South, only one school policy-community support variable is significant: days in session in the North and average time spent in guidance in the South. The very large difference in the constant term in the two equations is due to the absence of the days-in-session variable in the southern equation.

of class and home environment on scholastic achievement among pairs of genetically equivalent identical twins. (See Table 3.)

Although in Table 7 the corrected regression coefficients of all background variables are positive, in general the predicted effect of social class and family environment on the difference in scores at grade one and grade twelve is ambiguous. This is because the grade-one scores in

TABLE 7

Correction For Specification Bias Due to Omitted Initial Endowments in the Educational Production Function, Black Twelfth-Grade Students

Independent Variable (dependent variable is verbal achievement)	Regression Coefficients: At Grade Twelve[a] (1)	Regression Coefficients: At Grade One[b] (2)	Corrected Regression Coefficients[c] (3)
Reading material in the home	1.657	.348 (1.97)	1.029
Number of siblings (positive = few)	1.758	–	1.758
Parents' educational level	2.451	.884 (5.85)	.856
Family stability	.833	–	.833
Teacher's verbal ability score	1.041	–	1.041
Science lab facilities	.037	–	.037
Average time spent in guidance	1.480	–	1.480
Days in session	.203	–	.203

[a]From Table 6.

[b]t ratios are in parentheses; the coefficient of determination for the equation was .05.

[c]Column 3 = Column 1 – Column 2 × $b_{1,12}$, where $b_{1,12}$, the regression coefficient of A_1 in equation 7, is assumed to be

$$.5\left(\frac{\sigma_{12}}{\sigma_1}\right), \text{ or } 1.806.$$

equation (7) measure the initial level of learning and not the student's capacity to learn. There is evidence that among below-average achieving children with similar IQ scores in the elementary school grades, learning capacity on some tasks varies widely, and *inversely* with social class. That is, lower-class students with an IQ of 90 learn faster than upper-class students with equivalent IQ scores on tasks which do not depend on previous learning.[73] Thus it would not be anomalous to find a negative relationship in equation (7) between twelfth-grade scores and social background, at least for some background variables.

The main difficulty with the above method arises because of the social selectivity of the dropout phenomenon. The students tested in the fall of the twelfth grade are not the full complement of those who began in grade one. The fact that low achieving, low-social-class children are much more likely to drop out than other children, clearly biases downward the estimated coefficients of the social class variables in equation (8).[74]

Turning now to a second problem of specification bias, recall that equation (2) represented our reduced form. Yet a complete specification of the learning environment must include student attitudes. As mentioned earlier, these attitudes are represented in two ways: student self-concept and student sense of control over environment.[75] Measurements of these are added to the equation and the resulting estimates are presented in Table 8.

In the new equation, the structural parameters of school inputs change very little, which suggests that in this case the simultaneous-equation bias is relatively small. The attitude variables are powerfully related to achievement—the proportion of variance explained is almost doubled by their inclusion.

[73] Jensen [41]. Of course, much of school learning does depend on prior learning, but a major portion of it probably is not strictly cumulative once the rudimentary communications skills have been learned.

[74] The seriousness of this bias is difficult to determine, although experiments using a similar equation for grade-nine students (who have not had the opportunity to drop out) may suggest the order of magnitude of the bias. Of the four regression coefficients for the relevant social-background variables in an equation predicting ninth-grade achievement (rescaled to take account of the different units of measurement of the achievement variable), two are similar to those in the twelfth-grade equation and two are about 50 per cent larger than the downward biased estimates at grade twelve.

[75] The measurement of these variables is described in the appendix.

7 CONCLUSION: THE EFFECTS OF SCHOOLING

IMPERFECT measurement, limited exposure to the educational environment, and our fundamental ignorance of how children learn establish the presumption that estimated effects of different schools upon scholastic

TABLE 8

Educational Production Function with Student Attitudes Measured, Black Male Twelfth-Grade Students

Independent Variable (dependent variable is verbal achievement)	Regression Coefficient (*t* in parentheses)	Beta
1. Reading material in the home	0.4982 (0.7169)	0.0212
2. Number of siblings (positive = few)	1.5287 (3.8885)	0.1087
3. Parents' educational level	1.8746 (3.6768)	0.1088
4. Family stability	0.3818 (0.8489)	0.0228
5. Science lab facilities	0.0355 (1.9383)	0.0552
6. Days in session	0.1814 (1.8571)	0.0519
7. Teacher's verbal-ability score	1.1100 (6.4133)	0.1966
8. Average time spent in guidance	1.7747 (3.0644)	0.0964
9. Student's control of environment	4.4059 (8.2159)	0.2334
10. Student's self-concept	4.2721 (7.4439)	0.2108
Constant:	-12.1269 (-0.6949)	
R^2 :	0.3036	
$\lvert X'X \rvert$:	0.3764	
Number of observations:	1,000	

TABLE A.1

Full List of Variables Used

Dependent variable:	
	Verbal score
Nonschool environment:	
	Consumer durables in the home
	Family stability
	Foreign language at home
	Number of siblings
	Parents' educational level
	Preschool attendance
	Reading material in the home
	Student urbanism of background
General school environment:	
	Accelerated curriculum
	Amount of homework
	Average time spent in guidance
	Comprensiveness of curriculum
	Days in session
	Extracurricular activities
	Freedom of movement between tracks
	Length of academic day
	Number of foreign language courses
	Number of mathematics courses
	Number of twelfth-grade students in school
	Promotion of slow learners
	Proportion of students transferring in and out
	Teacher turnover
	Tracking
Teacher quality:	
	Degree received (teacher)
	Experience (teacher)
	Localism (teacher)
	Number of absences (teacher)
	Quality of college attended by teacher
	Salary (teacher)
	Teacher's socio-economic status
	Teacher's verbal-ability score
Teacher quantity:	
	Total pupils in school / total teachers in school

(continued)

Table A.1 (concluded)

School facilities:
Science laboratory facilities
Volumes per student in the school library
Student attitudes:
Student self-concept
Student sense of control of environment

achievement will be quite limited. Our equation (Table 6) suggests that, for the individuals represented by our sample, a uniform improvement of 10 per cent in *all* school inputs (in the neighborhood of the mean) would raise achievement by 5.7 per cent. (This cannot be construed as the effect of a uniform improvement in the school environment, as we have not measured some of the important school inputs.) Put somewhat differently, the difference in achievement between students in schools with inputs one standard deviation below the mean for our sample, compared with students in schools one standard deviation above the mean, is slightly over two-thirds of a standard deviation on our achievement scale.[76]

Given the limited nature of the sample and the inadequate opportunity to explore the available data, I will refrain from generalizing from these initial encouraging results. We are still a long way from estimating a satisfactory educational production function. However, we have successfully identified a number of school inputs which do seem to affect learning. Further studies of the educational production function may contribute to making schools more nearly equal for those now deprived of a constructive learning environment, and more effective for all children.

[76] This is roughly equivalent to two years of scholastic progress, using the performance of white urban Northeastern children as the norm. For the purposes of these calculations, the variable "length of school year" is considered a community variable, not a school input.

TABLE A.2

Means, Standard Deviations and Zero-Order Correlations Among Variables Used in Estimates

Variable[a]	Mean	Standard Deviation
Dependent variable		
Verbal achievement scale score[b]	49.2202	14.4512
Home environment [c]		
Reading material in the home	-0.1091	0.6159
Number of siblings (positive = few)	-0.3334	1.0275
Family stability	-0.1691	0.8645
Parents' educational level	-0.1672	0.8389
School environment		
Teacher's verbal-ability score	21.2211	2.5593
Science lab facilities (index)[d]	89.4083	22.4557
Average time spent in guidance	1.8528	0.7847
Number of days in session	179.8984	4.1359
Size of the senior class	264.3718	212.7663
Student attitudes		
Sense of control of environment[e]	-0.1265	0.7654
Self concept[f]	0.0460	0.7132

[a]Further definition of these variables, as well as the survey instruments on which they were based, is available in J. S. Coleman *et al.*, *Equality of Educational Opportunity* [13], Vol. II.

[b]The verbal-ability score is based on the School and College Ability test scores of the Educational Testing Service.

[c]The home environment and student attitude variables have been normalized to mean = 0 and standard deviation = 1 for the national sample taken as a whole.

[d]Range = 0–99. A score of 33, 66, or 99 indicates that the school has one, two, or all of the following types of laboratories: biology, chemistry, and physics.

[e]The sense of control variable is based on the student agreement or disagreement with three statements: "Good luck is more important than hard work for success; Every time I try to get ahead, something or somebody stops me;" and "People like me don't have much of a chance to be successful in life."

[f]The self-concept variable is based on the student's responses to the following items: (1) "How bright do you think you are in comparison with the other students in your grade?" (2) "Sometimes I feel that I just can't learn" (agree – disagree); (3) "I would do better in school work if teachers didn't go so fast" (agree – disagree).

TABLE A.3

Zero Order Correlations Among Variables

Variables	1	2	3	4	5	6	7	8	9	10	11	12
1. Verbal achievement scale score	1.000	.220	.225	.044	.261	.308	.188	.254	.153	.262	.350	.282
2. Reading material in the home		1.000	.178	.107	.363	.204	.152	.218	.106	.216	.161	.117
3. Number of siblings			1.000	-.049	.263	.152	.122	.134	.108	.173	.089	.042
4. Family stability				1.000	-.020	-.019	.001	-.010	.009	-.018	.094	.044
5. Parents' educational level					1.000	.198	.136	.174	.050	.205	.112	.115
6. Teacher's verbal-ability score						1.000	.280	.453	.179	.506	.054	-.053
7. Science lab facilities							1.000	.295	.170	.263	.078	-.039
8. Average time spent in guidance								1.000	.287	.522	.063	-.071
9. Number of days in session									1.000	.304	.058	-.026
10. Size of the senior class										1.000	.070	-.065
11. Sense of control of environment											1.000	-.309
12. Self-concept												1.000

APPENDIX

SUPPLEMENTARY STATISTICAL INFORMATION

The final specifications of the educational production functions were based on analysis of data on a much larger number of variables. These are listed in Table A.1. The means and standard deviations of the variables appearing in the final specifications appear in Table A.2., and the zero order correlations among these variables appear in Table A.3. All

TABLE A.4

Estimated Production Function for Samples of Northern U.S. Twelfth-Grade Students

Independent Variable[a]	Regression Coefficient (*t* in parentheses)	Beta
1. Reading material in the home	1.279 (1.601)	.052
2. Number of siblings (positive = few)	1.660 (3.700)	.116
3. Parents' educational level	2.655 (4.626)	.151
4. Family stability	.899 (1.675)	.051
5. Teacher's verbal-ability score	.721 (3.193)	.097
6. Science lab facilities	.059 (2.137)	.067
7. Days in session	.189 (1.971)	.062

Constant:	-2.585 (-0.1462)
R^2:	.090
$\lvert X'X \rvert$:	.730
Number of observations:	1,000

[a]Dependent variable is verbal achievement.

of the basic data in these appendix tables is based on the Equality of Education Survey of the U. S. Office of Education. The correlation coefficients, means and standard deviations are from J. S. Coleman [13]. The form and specification of the educational production function which is developed in the text and estimated for a national sample of 1,000 students was tested on different samples of 1,000 students each in the North and the South. The results of this estimation appear in Tables A.4, and A.5.

TABLE A.5

Estimated Production Function for Samples of Northern U.S. Twelfth-Grade Students

Independent Variable[a]	Regression Coefficient (t in parentheses)	Beta
1. Reading material in the home	1.841 (2.629)	.083
2. Number of siblings (positive = few)	1.794 (4.438)	.135
3. Parents' educational level	2.185 (4.181)	.132
4. Family stability	.823 (1.858)	.053
5. Teacher's verbal-ability score	1.097 (6.593)	.210
6. Science lab facilities	.027 (1.724)	.052
7. Average time spent in guidance	2.017 (3.266)	.102

Constant:	20.373 (6.247)
R^2:	.1961
$\lvert X'X \rvert$:	.519
Number of observations:	1,000

[a]Dependent variable is verbal achievement.

REFERENCES

1. Anastasi, Anne, "Intelligence and Family Size," *Psychological Bulletin,* vol. 53, no. 3 (August, 1956), 187–209.
2. Armor, David J., "School Effects on Negro and White Achievement: A Re-Examination of the USOE Data," 1968 (mimeo.).
3. Astin, A. W., "Undergraduate Achievement and Institutional Excellence," *Science,* vol. 161, August 16, 1968, 661–68.
4. Bereiter, C., *et al.*, "An Academically-Oriented Pre-School for Culturally Deprived Children," Paper presented at AERA meeting, Chicago, February 1965.
5. Bloom, Benjamin S., *Stability and Change in Human Characteristics*, New York, 1964.
6. Bowles, Samuel S., "Towards Equality of Educational Opportunity?" *Harvard Educational Review,* vol. 38, no. 1, (1968), 89–99.
7. Bowles, Samuel S., and Henry M. Levin, "More on Multicollinearity and the Effectiveness of Schools," *The Journal of Human Resources,* vol. III, no. 3, (Summer) 1968, 393–400.
8. ————, The Determinants of Scholastic Achievement—An Appraisal of Some Recent Evidence," *The Journal of Human Resources,* vol. III, no. 1, (Winter) 1968.
9. Burkhead, Jesse, Thomas G. Fox, and John W. Holland, *Input and Output in Large-City High Schools*, Education in Large Cities Series. Syracuse, New York, 1967.
10. Burt, C., "The Evidence for the Concept of Intelligence," *British Journal of Education,* vol. 25, (1945), 158–77.
11. Chausow, H. M., "The Organization of Learning Experience to Achieve More Effectively the Objective of Critical Thinking in the General Social Sciences Courses at the Junior College Level," unpublished Ph.D. dissertation, University of Chicago, 1955.
12. Cohen, J., "The Factorial Structure of the WISC at Ages 7–6, 10–6, and 13–6," *Journal of Consultative Psychology,* vol. 23, (1959), 285–99.
13. Coleman, James S., *et al., Equality of Educational Opportunity,* 2 vols. Washington, D. C., 1966.

14. Coleman, W., and E. E. Cureton, "Intelligence and Achievement," *Educational and Psychological Measurement*, vol. 14, (1954), 347–51.
15. Council of Economic Advisers, (*Annual Report*), Washington, D. C., 1967.
16. Dave, R. H., "The Identification and Measurement of Environmental Process Variables That Are Related to Educational Achievement," unpublished Ph.D. dissertation, University of Chicago, 1963.
17. Davis, A., and J. Dollard, *Children of Bondage*, American Council on education, Washington, D.C., 1940.
18. Deutsch, M., "The Disadvantaged Child and the Learning Process," in A. H. Passow (ed.), *Education in Depressed Areas*, New York, 1963.
19. Duncan, O. D., "Achievement and Ability," *The Eugenics Quarterly*, 1968.
20. Duncanson, J. P., "Learning and Measured Abilities," *Journal of Educational Psychology*, vol. 57, (1966), 220–29.
21. Farrar, Donald E., and Robert R. Glauber, "Multicollinearity in Regression Analysis: The Problem Revisited," *Review of Economics and Statistics*, (February 1967), 98–99.
22. Flanagan, John C., *et al.*, *Project Talent*, Pittsburgh, vol. 1, 1962; vol. 2, 1964.
23. Freeman, F. N., H. H. Newman, and K. J. Holzinger, *Twins: A Study of Heredity and Environment*, Chicago, 1937.
24. Ghiselli, E. E., "The Measurement of Occupational Aptitude," *University of California Publications in Psychology*, vol. 8, no. 2, (1955).
25. Gray, Susan W., and R. A. Klaus, "An Experimental Preschool Program for Culturally Deprived Children," *Child Development*, vol. 36, (1965), 887–98.
26. Griliches, Zvi, "Notes on the Role of Education in Production Functions and Growth Accounting," this volume.
27. Gross, Morris, *Learning Readiness in Two Jewish Groups*, New York: Center for Urban Education, 1967.
28. Hanoch, G., "Personal Earnings and Investment in Schooling." unpublished Ph..D. dissertation, University of Chicago, 1965 (mimeo.).

29. Hansen, W. Lee, Burton A. Weisbrod, and William J. Scanlon, "Schooling and Earnings of Low Achievers," *American Economic Review,* vol. 60, no. 3, (June 1970).
30. Hanushek, Eric A., "The Education of Negroes and Whites," unpublished Ph.D. dissertation, Massachusetts Institute of Technology, 1968.
31. Harrell, Ruth F., Ella Woodyard, and A. I. Gates, *The Effect of Mothers' Diets on the Intelligence of the Offspring*, New York, 1955.
32. Hoch, Irving, "Estimation of Production Function Parameters Combining Time-Series and Cross-Section-Data," *Econometrica*, vol. 30, no. 1, (January 1962), 34–53.
33. ———, "Simultaneous Equation Bias in the Context of the Cobb-Douglas Production Function," *Econometrica*, vol. 26, no. 4, (October 1958), 566–78.
34. Hollingshead, A. B., *Elmtown's Youth*, New York, 1949.
35. Hunt, J. McV., *Intelligence and Experience*, New York, 1961.
36. Husén, Torsten (ed.), *International Study of Achievement in Mathematics,* vol. I, New York, 1967.
37. Jackson, J. D., R. D. Hess, and Virginia Shipman, "Communication Styles in Teachers: An Experiment," Paper presented at AERA meeting, Chicago, (February 1965).
38. Jensen, Arthur R., "Estimation of the Limits of Hereditability of Traits by Comparison of Monozygotic and Dizygotic Twins," *Proceedings of the National Academy of Sciences,* vol. 58, (1967), 149–57.
39. ———, "How Much Can We Boost IQ and Scholastic Achievement?" *Harvard Educational Review,* vol. 39, no. 1, (1969).
40. ———, "Patterns of Mental Ability and Socioeconomic Status," *Proceedings of the National Academy of Sciences,* (August 1968).
41. ———, "Social Class, Race, and Genetics: Implications for Education," *American Educational Research Journal*, vol. 5, no. 1, (January 1968).
42. Kain, John F., and Eric A. Hanushek, *On the Value of Equality of Educational Opportunity as a Guide to Public Policy*, Program on Regional and Urban Economics. Discussion Paper no. 36, Harvard University, May 1968.
43. Katzman, Martin T., *Distribution and Production in a Big City*

Elementary School System, Ph.D. dissertation, Yale University, 1967.

44. Kelley, T. L., *Interpretation of Educational Measurement*, Yonkers, New York, 1927.
45. Kiesling, Herbert J., "Educational Production Functions in New York State," 1968 (mimeo.).
46. ————, "Measuring a Local Government Service: A Study of School Districts in New York State," *The Review of Economics and Statistics,* vol. XLIX, no. 3, (August 1967), 356–67.
47. Kirk, S. A., *Early Education of the Mentally Retarded*, Urbana, Illinois, 1958.
48. Levin, Henry M., *Recruiting Teachers for Large City Schools,* Studies in Social Economics, Washington, D. C., June 1968.
49. Machlup, Fritz, *The Production and Distribution of Knowledge in the United States,* Princeton, New Jersey, 1962.
50. Marschak, Jacob, and William H. Andrews, Jr., "Random Simultaneous Equations and the Theory of Production," *Econometrica,* vol. 12, nos. 3–4, (July-October 1944), 143–205.
51. Massell, Benton F., "Elimination of Management Bias From Production Functions Fitted to Cross-Section Data: A Model and An Application to African Agriculture," *Econometrica*, vol. 35, nos. 3–4, (July-October 1967), 495–508.
52. McDill, Edward L., Edmund D. Meyers, Jr., and Leo C. Rigsby, "Institutional Effects on the Academic Behavior of High School Students," *Sociology of Education,* vol. 40, no. 3, (1967), 181-99.
53. Michelson, Stephan, *Incomes of Racial Minorities*, Studies in Social Economics, Washington, D. C., September 1968.
54. Nerlove, Marc, *The Estimation and Identification of Cobb-Douglas Production Functions*, 1967.
55. Olim, E. G., R. D. Hess, and Virginia Shipman, "Relationship Between Mothers' Abstract Language Style and Abstraction Styles of Urban Pre-School Children," Paper presented at the Midwest Psychological Association meeting, Chicago, April 1965.
56. Peaker, G. F., *et al.*, *Children in Their Primary Schools*, Report of the Central Advisory Council for Education (England), vol. II, London, 1967.
57. Peterson, R. A., and L. DeBord, "Educational Supportiveness of the Home and Academic Performance of Disadvantaged Boys,"

IMRID Behavioral Science Monograph, no. 3, Nashville, Tenn., 1966.

58. Rosenthal, Robert, and Lenore Jacobson, *Pygmalion in the Classroom*, New York, 1968.
59. Scannell, D. P., "Differential Prediction of Academic Success, from Achievement Test Scores," unpublished Ph.D. dissertation, State University of Iowa, 1958.
60. Stodolsky, Susan S., and Gerald Lesser, "Learning Patterns in the Disadvantaged," *Harvard Educational Review*, vol. 37, no. 4, (1967), 546–93.
61. Terrell, G., Kathryn Durkin, and M. Wiesley, "Social Class and the Nature of Incentive in Discrimination Learning," *Journal of Abnormal Social Psychology*, vol. 59, (1959), 270–72.
62. Theil, H., "Specification Errors and the Estimation of Economic Relationships," *Revue Institute Internationale de Statistique*, vol. 25, (January 3, 1957), 41–51.
63. U. S. Commission on Civil Rights, *Racial Isolation in the Public Schools*, Washington, D. C., 1967.
64. Vandenberg, S. G., "Hereditary Factors in Psychological Variables in Man, with a Special Emphasis on Cognition," in J. S. Spuhler (ed.), *Genetic Diversity and Human Behavior*, Chicago, 1967.
65. Warner, W. L., R. J. Havighurst, and M. B. Loeb, *Who Shall Be Educated?* New York, 1944.
66. Welch, Finis, "Measurement of the Quality of Schooling, The Economics of Education," *American Economic Review*, vol. LVI, no. 2, (May 1966), 379–92.
67. Weisbrod, Burton A., and Peter Karpoff, "Monetary Returns to College Education, Student Ability, and College Ability," *Review of Economics and Statistics*, vol. 50, no. 4, (November 1968), pp. 491–497.
68. Weiss, Randall Dunn, "The Effect of Education Upon the Earnings of Whites and Blacks: Experiments with Single Equation and Recursive Models," *Review of Economics and Statistics*, May, 1970.
69. Wilson, A. B., "Social Stratification and Academic Achievement," in A. H. Passow (ed.), *Education in Depressed Areas*, New York, 1963.
70. Wolf, R. M., "The Identification and Measurement of Environmental Process Variables Related to Intelligence," unpublished Ph.D.

dissertation, University of Chicago, 1963.
71. Wolfle, Dael, and Joseph Smith, "The Occupation Value of Education for Superior High School Graduates," *Journal of Higher Education,* (April 1956), 201–13.
72. Zigler, E., and J. DeLabry, "Concept Switching in Middle-Class, Lower-Class and Retarded Children," *Journal of Abnormal Social Psychology,* vol. 65, (1962), 267–73.

COMMENTS

JOHN E. BRANDL

OFFICE OF THE SECRETARY, DEPARTMENT OF HEALTH, EDUCATION, AND WELFARE

Sam Bowles has written a discerning, discouraging paper. He has explicitly shown the gap between what estimators of education production functions have been doing (mostly least-squares regressions of achievement or even income on a host of explanatory variables), and what they could be doing with the econometric tools now available. The paper hints at the desirability of a system of simultaneous relations over a uniequational model, cleverly manipulates and combines variables to lessen the difficulties associated with collinearity, mentions that production functions should recognize that "schools are multiproduct firms," makes a start at giving some theoretical content to education production functions, produces some impressive estimates, and, in general, shows the way toward more respectable econometric analysis in this field that has been characterized by shabby statistics.

The paper is discouraging to me, however, for it leaves little room for hope that estimations of education production functions from survey data can be very useful for policy purposes in the foreseeable future. The remainder of this comment consists of two parts: the first outlines my reasons for taking this pessimistic position, and the second offers some observations on policy making given such a state of affairs. As the lone bureaucrat participating in this conference I am looking for answers to the question, "How can education expenditures be allocated more efficiently now?"

NOTE: Dr. Brandl is now Director, School of Public Affairs, University of Minnesota.

WHY PROSPECTS ARE BLEAK

Consider Bowles' regression equation (p. 13)

$$\hat{A} = \hat{f}_0 + \hat{f}_1 x_1 + \hat{f}_2 x_2 + \ldots + \hat{f}_n x_n$$

where $\hat{A}$ is a measure of "output," such as an achievement test score, the x_i are "input" measures, and the $\hat{f}$'s are constants. This is the typical specification of education "production functions," and it is used by Bowles despite his own cogent and compelling reservations. The implications of such a relationship for a policy-maker are disconcerting. As Bowles notes, optimum conditions for resource allocation are

$$\frac{\partial A}{\partial X_j} \bigg/ \frac{\partial A}{\partial X_k} = \frac{\hat{f}_j}{\hat{f}_k} = \frac{p_j}{p_k}$$

(where p_i is the price of x_i) or, alternatively, $\frac{\hat{f}_j}{p_j} = \frac{\hat{f}_k}{p_k}$, or achievement gain per marginal dollar expended should be identical for all factors.

The $\hat{f}$'s are constant by definition, and for any single decision-making body in this country (probably including the federal government since its contribution to education is relatively small) the p's probably are, too. Policy-makers are being told, then, that their job is to determine the highest ratio of $\hat{f}$ to p for any variable over which they have control and to expend all available funds in that direction. That conclusion is enough to disenchant any educator or bureaucrat—many of whom are already wary of economists that they think intend to show them up.

The following are some reasons why I believe such incredulity to be justified.

1. Surely the "production function" is misspecified. The argument against linear production functions already mentioned is at least as strong on a priori grounds as that suggested by Bowles against the Cobb-Douglas.[1]

[1] The Cobb-Douglas function would require, as Bowles mentions, "that an increase in the quality of teachers be more effective on children of well-educated parents than on the children of illiterate parents," i.e., that $\frac{\delta^2 A}{\delta x_j \delta x_k} > 0$. See Bowles (p. 19, note 9).

2. A variation on this specification theme is that we know and desire that our schools produce more than ability to achieve high scores on standardized achievement tests. Failure to include others of the joint outputs gives no indication of the effect of an input change on those other outputs. More importantly, it means that the regression coefficients do not even necessarily indicate the most efficient way to produce an impact on the single measure chosen, since the relative weightings of the several outputs in the preference functions of administrators are not known.

This leads to a third objection which I consider to be a critical blow to the usefulness of production functions of this sort. But to make myself clear, let me assume away the econometric problems alluded to—specification, simultaneous equations, multicollinearity (and others such as the absence of good time series data).

3. Even if we could overcome these statistical problems, the meaning of our estimates would be in doubt. Local, state, and federal school administrators and decision makers are maximizing neither "achievement" nor any other known combination of outputs, or rather, it is likely that different decision makers weigh the several outputs in different ways. Survey or longitudinal data which aggregate information resulting from the maximization of dozens, hundreds, or thousands of different objective functions will not yield an answer to the question posed in the beginning of this comment: "How can education expenditures be allocated more efficiently?"

WHERE DO WE GO FROM HERE?

If we had a well-defined theory giving us grounds for choosing particular variables as the inputs and outputs to be included,[2] and which provided guidance in the choice of functional form, then perhaps at least negative results of estimated production functions could be used to prod an embarrassed education community into doing better. That is, indications that marginal products of some inputs approach zero might be a goad to improve. As it is, tentative results showing little improvement

[2] Presumably the outputs could be chosen on either consumption or investment grounds. At present we are not able to relate our test measures either to the fun of being in school or to the financial gains of having been educated.

on achievement tests after injection of Head Start or Title I (of the Elementary and Secondary Education Act) funds[3] are often shrugged off by educators on the grounds that that's not what they are trying to do.

We are, then, in a pre-Newtonian (or perhaps even pre-Ptolemaic!) state where we not only lack theory, but have at least some grounds for believing that the best technology is not widely used.[4]

My personal reaction to this state of affairs can be summarized as follows:

a. Estimation of efficient technological relationships might be more readily accomplished through examination of exemplary and demonstration programs than through scrutiny of data from large national surveys—this not only because the objectives of such programs are more easily determined than are the objectives of a conglomeration of systems, but also because the statistical problems of managing data are not as great when one can influence selection of control and experimental groups, variables on which data will be collected, types of tests administered, and so on.

b. We economists tend to find more maximizing than exists. Just as the problem of X-efficiency[5] in firms has been neglected until very recently (with economists assuming that firms operate on their efficiency frontiers) our usual approach for estimating education production functions ignores not only the existence of joint products, but the possibility that our schools may not be efficient users of resources. Thus, survey data (such as that collected by the Coleman Commission, or the reams of material obtained each year through the survey of school districts receiving Title I ESEA funds) can be of more value for *describing* the present state of education, than for *prescribing* what might be done.

c. Attempts should be made to determine what it is that school

[3] See *Evaluation of Title I*, Office of Program Planning and Evaluation, Office of Education, Department of Health, Education, and Welfare, Washington, D. C., 1968.

[4] *Ibid.*, for unimpressive results of the present school system. Scattered examples of impressive results indicate the possibility that superior technology may exist but, for whatever reason, not be widely disseminated. See *A Study of Selected Exemplary Programs for the Education of Disadvantaged Children*, American Institute for Research in the Behavioral Sciences, Palo Alto, California, September 1968.

[5] See H. Leibenstein, "X-Efficiency vs. Allocative Efficiency," *American Economic Review*, June 1966.

authorities are trying to accomplish, first, in order to find out whether they are attaining their objectives, and second, so that the questions of whether to change those objectives and provide incentives to do so can be considered.

d. In addition to fostering research to find out what works and how to get people to use it, we should recognize that our current ignorance has implications for how we allocate money to existing programs. For example, in the field of higher education, there is a growing consensus (both inside the government and out) that federal assistance to students is preferable to the government's buying *inputs* (such as buildings, teaching paraphernalia, or teachers). That is, for the time being we cannot relate pompons per cheerleader to lifetime income or graduate record exam score, so perhaps assisting students directly—and letting them choose where (and to some extent, how) to spend the money—has some advantages.

Sam Bowles, cosmetologist,[6] has beautified considerably what was a most unattractive client. Unfortunately, she is still not very helpful around the house.

JOHN C. HAUSE
UNIVERSITY OF MINNESOTA

In the following comment Bowles' interesting, exploratory paper will be discussed under three headings: first, some important aspects of his general analytic framework; second, several specific problems arising in the execution of the study; and finally, a few of the main issues the study raises for future research.

I. GENERAL FRAMEWORK OF THE STUDY

Bowles sets for himself the task of devising a systematic and sensible framework for studying the relationship between inputs and outputs of schools. In addition, he presents some initial attempts to estimate

[6] See Bowles (p. 26).

crude structural parameters of relationships implied by this framework. The motivation of the study appears to be essentially normative—presumably to aid in improving allocation in the educational sector. It is not concerned with how or why school administrators currently allocate educational resources as they do.[1] Indeed, Bowles explicitly assumes that "they do not select or alter school inputs with a mind to optimizing any well-defined function of school outputs." As the author recognizes, this implies that the estimated dependence of educational outputs on inputs may significantly understate what could be achieved if the inputs were allocated in an efficient way.[2] One might perhaps examine in greater detail schools with large positive residuals from regressions in hopes of determining additional factors that "explained" why these schools were especially productive. But cross-sectional studies of the average observed relationships between inputs and outputs do not appear to be very promising as a way of getting at optimal techniques in the current regime. However, Bowles considers that the most serious difficulty is the lack of a theory of learning that would aid in specifying the production function. I shall discuss the implications of this problem.

One important omission in Bowles' general analysis is a careful discussion of the appropriate level of aggregation for carrying out the study of the production of education, especially if the results are intended to held guide the decisions of school administrators. As the author recognizes, the lengthy process of educating a child is very complex, and the measurable inputs applied by school administrators vary greatly in magnitude and importance during this process. He stresses the important role of value added analysis for appropriately imputing the returns from education, but does not adequately emphasize the main methodological issues that depend on the use of a value added framework. Consider a

[1] One might argue that positive economic analysis of resource allocation in the public sector is at least as deficient as relevant normative analysis. Modification of existing incentives and constraints in this sector may at times be as important as presenting confused administrators with better normative theories to aid them in their decisions.

[2] Bowles provides some indirect evidence which suggests current inefficiencies. For the educational outputs he is concerned with, instructional costs per pupil are less strongly related to output than specific characteristics of the inputs, which in turn can also "explain" much of the variance of expenditures per pupil.

child who has been "in process" in the educational system for a number of years. Is it possible to measure his current status by means of achievement, aptitude, and interest tests and thereby determine the additional inputs that will efficiently help him achieve additional, well-defined educational objectives? Or is the educational process so interdependent that the full, detailed history of the child's previous educational experience critically affects the appropriate inputs for the next stage of development? The task of improving the technology of education is greatly simplified if extensive decomposition of the education production function is possible for a least two reasons. First, it permits a degree of decentralization and the assignment of limited, but well-defined responsibilities to teachers and administrators who are directly concerned with only a small part of the whole educational process. Second, if such decomposition of the educational process is feasible, it may help to identify certain kinds of achievement and attitudes which are exceptionally important for subsequent educational development. It seems possible that a significant increase in educational expenditures on such key elements would yield high returns if they could be identified.

The extent to which such temporal disaggregation of the educational production function is useful is essentially the same question as the extent to which detailed value added analysis is feasible. This is currently an open question. Bowles cites an article by Jensen that indicates possible pitfalls in trying to specify the current status of a child that would be appropriate for a value added analysis. Jensen mentions some evidence that students from lower social-economic backgrounds with a given IQ learn faster than students from higher backgrounds with the same IQ when confronted with tasks that are not dependent on former learning. Thus IQ information alone might lead to inappropriate groupings of young pupils for some purposes.

Even if it is impossible to disaggregate the detailed education production function very far, it is not clear how much normative guidance can be obtained from a highly aggregated model. The statistical experiments an economist can carry out depend, of course, on the data available to him. Still, an explicit discussion of the appropriate level of aggregation might be useful to other people studying these problems.

II. SEVERAL TECHNICAL PROBLEMS.

Although Bowles considers the empirical calculations reported in the study to be preliminary, several issues deserve comment. A key difficulty stems from the lack of explicit variables for controlling initial ability (or achievement). The author acknowledges that such a variable would be desirable, but it is not available in his cross-section data. As a partial adjustment, he adjusts home environment regression coefficients for twelfth graders by the corresponding coefficients for first graders. The motivation for this procedure is not very clear, and the modification in any case seems to be of dubious value. This adjustment would seem to be relevant for discussing the following question. Given the influence that home environment has already exerted on a child by the time he has entered the first grade, how much additional influence does it have on his achievement test score by the time he is in the twelfth grade? It is not obvious what issue is clarified by such an imputation even if it were legitimate. The procedure does not seem appropriate because the home environment variables are surely correlated with omitted variables (including genetic factors) that are determinants of first grade "ability" or achievement. Strong (and implausible) assumptions are required to justify interpreting the adjusted coefficient as a measure of the net *additional* influence of home environment after the child has entered the first grade.

The limited predictive power of Bowles' initial regressions is of course due primarily to the unavailability of control for initial ability. Nevertheless, the relationships of central importance for this study are those between school environment and education output variables. It is not clear how much the coefficients of these variables will be biased by inability to control for initial ability and this issue is more important than the values of the coefficients of determination obtained in the reported regressions. It seems quite possible that inefficiencies in the current allocation of educational resources may be empirically the more important cause of understating the size these coefficients would have if the resources were appropriately organized.

The general variables that seem appropriate determinants of the educational production function are not directly observable. This raises the problem of the appropriate way of aggregating less satisfactory proxy variables into indexes representing the general factors. The difficulty with including all the specific factors is that multicollinearity makes the estimates of the individual coefficients unstable and difficult to interpret. Bowles does not attempt to resolve this issue. He mentions two alternatives—selecting a single proxy variable to represent each general variable or adopting some formal statistical procedure such as principal components analysis—and opts for the former in this study. The following very simple procedure might be more appropriate than either of these suggestions. Associate the specific variables that seem to be reasonable proxies for a general factor with that factor, run an ordinary multiple regression including all these specific variables, and test only groups of coefficients, where each group includes the specific variables associated with a particular general factor. The test corresponding to the usual *t*-test for a coefficient being significantly different from zero is to test the vector of specific variable coefficients corresponding to a general factor to see whether it differs significantly from a zero vector. Multicollinearity will often lead to imprecise estimates of the coefficients of some specific variables, but it is presumably the influence of the general factor that is of interest. This procedure amounts to accepting the linear function determined by the coefficients of each group of specific variables as an appropriate index serving as a proxy for the general factor.

III. SOME IMPLICATIONS AND PROBLEMS FOR FUTURE RESEARCH.

Bowles recognizes that such educational "outputs" as achievement scores in tests do not represent the ultimate outputs that one hopes the educational system will generate. Specific levels of proficiency may indeed provide minimal standards of achievement which are useful tools for school administrators. Bowles discusses briefly the importance of trying to relate the educational process to such outputs as earnings capacity. The implicit conclusion appears to be that it is necessary to analyze how the more specific, readily measured educational outputs are related to these more important outputs. This raises the question whether the analysis of the relationship between the "ultimate" outputs and inputs allocated by school administrators should attempt to establish a direct relationship or whether these variables should be related only by intermediate output variables such as test scores. At this stage of research into the economics of education, this is an unresolved issue. Certain inputs may generate so many joint products of intermediate educational output that the appropriate imputation of ultimate returns to school-administered inputs will be an unmanageable task if the analysis must proceed through the mechanism of the intermediate level outputs of education. At least this possibility should be considered before any major attempt is made to analyze the ultimate outputs in an educational production function framework.

A second important issue for future research was raised in the discussion in the first part of this comment: the levels of aggregation that will be most fruitful for improving the technology of education. The magnitude of sub-optimal decisions currently made by administrators may seriously limit the relevance of statistical analyses based on typical, current practices for improving current technology.