On Equality
of Educational
Opportunity

PAPERS DERIVING FROM THE HARVARD UNIVERSITY
FACULTY SEMINAR ON THE COLEMAN REPORT

EDITED BY

Frederick Mosteller
&
Daniel P. Moynihan

Contents

IMPLICATIONS FOR THE FUTURE

APPENDICES: THE PROBLEM OF MEASUREMENT: Survey & Model

THE COLEMAN REPORT

3

On the Value of

Equality of

Educational

Opportunity as a

Guide to Public Policy*

ERIC A. HANUSHEK & JOHN F. KAIN[1]

Equality of Educational Opportunity has been part of the public record since July, 1966.[2] The best known "finding" of the *Report* is that quantity and quality of school inputs (facilities, curriculum, and personnel) have little or no bearing on achievement; home environment and the student's peers are what really count.[3] Obviously, such a finding has far-reaching implications for educational policy. At the very least, it raises serious questions about the efficacy of the billions of dollars now spent on public education. Yet in our opinion, serious doubts must be raised about this and several other "findings" attributed to the *Report*.[4] These doubts result both from the methods of empirical analysis and their interpretation. While we are not the first to raise questions about the *Report*'s analysis, we feel that the subject is both important and complex enough to merit further discus-

sion.[5] Our discussion of the analysis presented in *Equality of Educational Opportunity* and associated interpretative problems is preceded by a consideration of the more fundamental area of overall research strategy. Past reviews of the *Report* have concentrated on assessing the analysis and findings within the framework defined by the *Report*'s authors. We are interested, as well, in the merits of the work within the framework defined by the original Congressional directive.

I. ALTERNATIVE
RESEARCH STRATEGIES
—AN INTERDEPENDENT
DECISION

We contend that the authors of *Equality of Educational Opportunity* made a fundamental error in confusing a responsibility for fact-finding with a mandate to carry out basic research on the educational production process. This had repercussions on all aspects of the study because the pressures of time, knowledge, and available resources were magnified by attempting the broader study in conjunction with the required fact-finding mission. The result was a failure to provide the information requested by Congress.

The *Report* was the response of the Office of Education to Section 402 of the Civil Rights Act of 1964 which stated:

> The Commissioner [of Education] shall conduct a survey and make a report to the President and the Congress, within two years of the enactment of this title, concerning the lack of availability of equal educational opportunities for individuals by reason of race, color, religion, or national origin in public educational institutions at all levels in the United States, its territories and possessions, and the District of Columbia.

There is little doubt that Congress wished to obtain an authoritative answer to the question of whether minorities were being discriminated against in the provision of public education.[6] However, this question is not as simple as it first appears. At least two possible definitions of equality of educational opportunity come to mind: (1) equality of resources or school inputs and (2) equality of achievement or output of the educational process.[7] All subsequent decisions about the research design and data needs depend on which definition is selected.

Measurement of the educational resources provided students of each minority group is the backbone of an analysis of school input inequality. Resources can be measured in either real terms (quantities of appropriately

weighted homogeneous inputs) or money terms (per pupil expenditures).[8]
The crucial feature of such a survey is obtaining representative samples
from which population inferences can be made. There are some difficult
conceptual and measurement problems of the input method, but they are not
insurmountable. In fact, a major strength of this approach is simplicity (a
significant consideration given the time constraint on the study). Addition-
ally, such a study of input inequalities provides immediately usable infor-
mation for initial legislative or judicial action.

For an output definition of equality of educational opportunity, the
focus of the data collection should be on achievement levels of a representa-
tive sample of population groups. Again it is essential to be able to make
inferences about the entire population.[9] If large inequalities in the average
level of such income-related output measures are found to exist among
groups in society, the policy objectives are quite clear, even if the exact
means of achieving these objectives are not.

While it is conceptually a rather simple matter to equate educational
inputs, it is much more difficult to devise policies that will equalize expected
output. This difficulty may have been what motivated the Office of Educa-
tion to commission a very broad and ambitious program of basic research on
the educational production process. This larger analysis took the form of
estimating a multivariate statistical model relating student and school char-
acteristics to achievement. As is discussed in Section II of this paper, this
effort was unsuccessful. The relevant issue at this point is the way in which
the decision to undertake this ambitious research affected the evaluation of
equality of educational opportunity.

Analysis of the educational production process requires the collection
of more precise and detailed input and output data than those required for
both the input and output investigations combined. These data should in-
clude information on both outside school factors (socioeconomic status,
family attitudes, community environment, and similar factors) and past
school inputs (longitudinal data) in addition to the levels of current school
inputs (the survey of input equality) and current achievement (the survey
of output equality). For a study of the educational production process, it is
more important to obtain wide variation in educational practice and experi-
ence than to have a representative sample of the population of schools or
students. These differences can be critical when there are time and resource
constraints on the study.

We contend the Office of Education should have been less ambitious in
its investigation and limited itself in the short time period available to fact-
finding. We would be the last to argue against the need for more basic
research on the production of educational achievement or to insist on a nar-
row interpretation of Congressional intent; however, we ask whether the
immediate needs of public policy would not have been better served by a
careful and exacting determination of the narrower question of inequality in

the provision of educational resources—a question about which consider-
able controversy remains.[10]

Even the most permissive interpretation of the Congressional directive
could not relieve the authors of the *Report* of the requirement for a system-
atic survey of equality in the provision of educational resources. They could
do more, but not less. The clear danger of doing more was that the diver-
gent data requirements and the different areas of emphasis demanded by the
three research strategies (the input survey, the output survey, and the basic
research on the educational process) would prevent them from providing an
authoritative answer to any of the three questions. It appears that this is
precisely what happened. Confronted with a restricted choice among three
strategies, the authors chose all three. Had they succeeded in providing au-
thoritative answers about inequalities in educational inputs, we would have
had no quarrel with their decision to undertake research on the educational
process. However, in attempting to do all three, the authors of the *Report*
failed to provide convincing answers to the question of whether minority
groups are systematically discriminated against in the provision of educa-
tional resources.

♦ *The OE Survey*

The following discussion of the *Report*'s data base supports our con-
tention that the authors failed to provide satisfactory answers to the ques-
tions concerning equality of educational resources. Moreover, this section
is necessary for our discussion of the conceptual model and statistical meth-
ods used by the *Report*'s authors in their research into the educational proc-
ess. This data base is referred to as the *OE Survey* to differentiate it from
the analysis presented in *Equality of Educational Opportunity*.

The basic sampling units used in the *OE Survey* were elementary and
secondary schools attended by seven broad ethnic groups: whites, Negroes,
Oriental Americans, Indian Americans, Mexican Americans, Puerto Ri-
cans, and "others." High schools included in the *OE Survey* were selected
by a stratified probability sampling technique which insured that schools
attended by minorities were overrepresented. Elementary and junior high
schools were sampled on a probability basis depending on the percentage of
their students going to the secondary schools included in the final sample.
The sample size was set originally at 900,000 students, but nonresponses
reduced the usable sample to approximately 569,000 students. These stu-
dents, divided among grades 1, 3, 6, 9, and 12, were given ability and
achievement tests and completed a questionnaire concerning family back-
ground and attitudes. In addition, data were gathered from the teachers,
principals, and school-system superintendents for the 3155 sample schools.
Teachers completed an optional verbal facility test and a questionnaire on
their personal histories, educational backgrounds, attitudes, and character-

istics of their classes and schools. Principals and superintendents supplied information about their backgrounds and attitudes and about school facilities in their school or district.

There are several substantive problems with the *OE Survey*. Throughout the *Report*'s presentation, the reader is lulled into a false sense of security by the seemingly generous sample size (569,000 students). But, when it comes to school facilities, the relevant sample size is the number of schools, not the number of students. This reduction in effective sample size results from a failure to obtain school input data pertaining to individual students. Although the school sample is still quite large, it is reduced considerably if stratification is necessary (by grade, race, region, urban/rural). For example, the 12th-grade sample for the metropolitan South included data on only 78 schools, and only four of these had between 10 percent and 75 percent nonwhite students.[11] One is hesitant to make inferences, especially as concerns the effects of integration from an analysis of such small samples.

Nonresponse problems (which were glossed over in the *Report*) are another serious weakness of the *OE Survey*. Refusal to participate and faulty responses meant that 41 percent of the 1170 high schools included in the original random sample could not be included in the analysis. Similarly, only 74 percent of the sample of feeder schools were included in the final analysis. It is obvious from these statistics that extreme care must be exercised in making inferences about the population because analysis of the *OE Survey* data could be seriously misleading if this nonresponse were systematic. As mentioned previously, systematic nonresponse would be most serious in the case of the narrower questions relating to inequality in educational inputs or outputs, especially if the reasons for nonresponse were related to a "sensitivity" about real or believed inequalities. In fact, there are indications of such systematic nonresponse. Several large Northern central cities, where there has been considerable controversy about school discrimination, refused to cooperate.

Nonresponse to specific questions also presents serious problems. Analyses of the raw data by one of the present authors indicate that many questionnaire items are unusable because of nonresponse problems.[12] High rates of nonresponse are particularly characteristic of emotionally sensitive questions. For example, the principal's questionnaire includes three questions about the principal's attitudes concerning racial composition of the faculty, assuming student bodies of three different racial mixes. In a sample of about 300 *OE Survey* elementary schools in the northeast region, over one third of the principals failed to answer one or more of these questions. Substantial errors could be introduced by such internal nonresponse.[13]

Lying, exaggeration, and faulty responses to particular questions are a problem in any survey. However, the potential for error seems especially

large for the *OE Survey*. Questions requiring numerical answers, such as school enrollment, to be coded for mechanical scoring provide the most dramatic examples of faulty response. These numerical answers entered into such traditional policy variables as class size, per-pupil expenditure, and volumes per student in the library. Cross-checking of these questions uncovered a considerable number of obvious miscodings.[14] It would appear that the computer revolution has not yet been felt fully in the nation's elementary and secondary schools. Though the possibility of error may be greatest for questions requiring self-coding of numerical answers, the frequency of such errors raises serious doubts about the reliability of the *OE Survey* in general.[15]

Serious as the problems of sample reliability may be, the fundamental weaknesses of the questionnaires are even more harmful, especially in terms of measuring school facilities. The absence of questions with any qualitative bite is particularly noticeable. There are many questions that relate to the presence of particular attributes, but few that relate to their quality. This is true of the description of physical facilities such as laboratories, gymnasiums, and textbook availability, as well as features such as curriculum and specialized classes. Measures of school facilities are very insensitive and do a poor job of differentiating schools. While these problems are not restricted to school input data, the overriding interest in school effects and the uneven quality of the data (as discussed under "Contemporaneous Errors" in Section II) suggest that the problems resulting from the inadequate measurement of school inputs are most serious for the analysis.[16]

Related to these measurement problems, the questionnaires frequently seem to stop short of asking many logical and important questions. For example, neither expenditures per school nor information on school organization were collected, except in the crudest form, and information on the educational histories of sampled students was not obtained. This omission is particularly unfortunate given the multi-tier sampling design used in the *OE Survey*. A very large proportion of high-school and junior-high-school students necessarily attended lower-tier schools in the sample. Thus, even the sketchiest historical information would provide links between elementary, junior high, and high schools.

Many of these shortcomings appear to be the result of a decision to use the same questionnaires for all grade levels. For example, the principal's questionnaire could be answered by both an elementary-school and a high-school principal, and the teacher's questionnaire could be answered by a 1st-grade teacher and a high-school guidance counselor. To an even greater degree the weaknesses of the survey instruments appear attributable to a lack of careful prior specification of hypotheses. Again, the press of time, limitations of resources, the pathbreaking nature of the research, and an unwillingness to limit the scope of the investigation appear to be responsible

for these failings. Given these very real problems, the question remains whether the *Report*'s basic research finding, the no-school-effect conclusion, can be regarded as an adequate guide for public policy.

II. THE RELATIONSHIP

BETWEEN INPUTS AND

OUTPUTS —

SOME CONCEPTUAL

AND METHODOLOGICAL

ISSUES

In addition to the information problems discussed previously, the *Report* employs a number of questionable procedures. These factors lead us to seriously question whether the no-school-effect finding could have arisen from data inadequacies and analytical methods rather than any underlying behavioral relationship. At the very least, it is clear that the shortcomings of the data and the analytic methods used by the *Report*'s authors in studying the educational process raise serious problems of interpretation.

There are two major sets of issues central to any discussion of the *Report*'s findings on the relationship between educational inputs and outputs. These are: (1) the conceptual and statistical models of the education process used by the *Report*, and (2) the statistical methods employed in testing these models. These two areas of concern are inextricably interrelated, and both are strongly implicated in the *Report*'s no-school-effect conclusion. Thus, no rigorous segregation of them is attempted in the discussion that follows.

It is apparent that unfamiliarity with both the terminology and methodology used by the *Report* is responsible for much of the confusion surrounding it. Because these statistical concepts are crucial to understanding both the *Report*'s findings and our critique, we have made an effort to clarify some of these concepts and to provide definitions of technical terms.

♦ *The Conceptual Model*

Much of the appeal of the *Report*'s analysis arises because it seems to test empirically a conceptual model of the educational process that has wide acceptance.[17] Most researchers and educational policy-makers subscribe to a general conceptual model similar to that depicted by Equation 1. This

model states that the achievement of an individual (*ith*) student at time t [A_{it}] is some function [g] of his characteristics and those of *his* immediate family cumulative to time t [$F_i^{(t)}$]; of the characteristics of his peers [$P_i^{(t)}$]; of his initial endowment or innate ability [I_i]; and of the quantity and quality of educational (school) inputs consumed by him throughout his lifetime [$S_i^{(t)}$].

$$(1) \quad A_{it} = g(F_i^{(t)}, P_i^{(t)}, I_i, S_i^{(t)})$$

where

A_{it} = Vector of educational achievement of the *ith* student at time t,
$F_i^{(t)}$ = Vector of individual and family characteristics for the *ith* student cumulative to time t,
$P_i^{(t)}$ = Vector of student body characteristics (peer influences), i.e., socioeconomic and background characteristics of other students in the school cumulative to time t,
I_i = Vector of initial endowments of the *ith* individual,
$S_i^{(t)}$ = Vector of school inputs relevant to the *ith* student cumulative to t.

Two aspects of the conceptual model deserve further emphasis. Innate ability refers to a pure genetic input that should not be confused with I.Q. or any other common measure of ability. Though we do not know of any satisfactory method of measuring this elusive concept, its inclusion in the conceptual model of the educational process is nevertheless of the utmost importance. Its inclusion as a separate argument in the achievement function does not imply a fixed ability or predetermined growth theory of intelligence. In fact, the conceptual model depicted in Equation 1 hypothesizes a heredity-environment interaction.

Separate peer and family vectors are included in the model because they have different policy implications. The socioeconomic, cultural, and racial composition of schools can be modified; for example, many schemes for educational reform, such as educational parks, are attempts to achieve these ends. But to change or reduce the importance of the family background of the individual child requires much more radical surgery. It is necessary either to change the characteristics or attitudes of individual families or else to weaken their influence on the child. Both are difficult and highly controversial objectives.[18]

Although the conceptual model in Equation 1 was never presented in the *Report* in this form, it seems implied throughout the text.[19] More importantly, it appears that most readers of the *Report* accept something of this general nature as the model of the educational process tested in it. Actually, the statistical models employed in the *Report* differ considerably and in sys-

tematic ways from this conceptual model. Moreover, the divergences among them (i.e., errors in model specification) tend to bias the empirical findings toward showing negligible school effects.

♦ The Statistical Model

All of the *Report*'s empirical analyses assume that the conceptual model can be written in a linear form such as in Equation 2. This functional form requires that individual achievement be described by the linear addition of an array of n variables plus some random residual (error) e_i. (These n variables are elements of the four vectors included in the conceptual model.) Equation 2 hypothesizes that consistent behavioral relationships exist across individuals as represented by the parameters or slope coefficients, a_j.[20]

$$(2) \quad A_i = a_o + a_1 X_{i1} + a_2 X_{i2} + \ldots + a_n X_{in} + e_i$$

where

$$A_i = \text{achievement of the } ith \text{ student,}$$
$$X_{i1}, X_{i2}, \ldots, X_{in} = n \text{ explanatory variables corresponding to measurements of the arguments of Equation 1 for individual } i,$$
$$a_1, a_2, \ldots, a_n = \text{parameters of the educational process,}$$
$$e_i = \text{residual term or the portion of } A_i \text{ that cannot be explained by the explanatory variables.}$$

There are several ways in which this general statistical hypothesis can be tested. The procedure used in *Equality of Educational Opportunity* was to partition or allocate the variance (the average squared deviation of individual observations from the sample means) in achievement among sets of explanatory variables (roughly corresponding to the vectors in Equation 1) through a specialized analysis of variance procedure.[21] This method involves calculating the amount of explained variance of achievement resulting from inclusion of different sets of explanatory variables in a least squares regression equation, i.e., calculating increments to the variance "explained" by the regression (R^2).[22] The *Report*'s conclusions are based on an assessment of the amount of variance explained by collections of variables included in each of the vectors (family, peer, school).

♦ Analysis of Variance in a Complex World

The specific analysis of variance procedure used by the *Report* is completely straightforward as long as all of the explanatory variables are truly independent (are not themselves correlated).[23] This is not true of the ex-

planatory variables used in the *Report*. For example, higher-income suburbs pay their teachers more. Similarly, well-educated parents are more likely to be strongly motivated about their children's academic achievement, will be in a better position to help them with homework, and are likely to consider school quality in choosing a place to live.[24]

When explanatory variables are intercorrelated, interpretation of the analysis of variance becomes exceedingly difficult.[25] Only a part of the explained variance can be assigned uniquely to particular variables or vectors. "Interaction" terms that measure the joint contribution to explained variance of two or more variables or vectors become very important.[26] The analysis of variance procedure used in the *Report* treats these interaction terms in a very unusual manner. Explanatory variables are entered into the model in a predetermined order and only the increment to explained variance is assigned to each new variable or vector. Thus, the proportion of variance allocated to each variable or vector depends on the order in which they are entered. If two variables or vectors are highly intercorrelated, the first entered will be assigned both its unique contribution to explained variance and its jointly explained variance with all other variables or vectors (the interaction terms). Changing the order in which explanatory variables or vectors are entered changes the proportion of explained variance attributed to each. A clear understanding of this characteristic of the statistical technique is crucial to evaluating the findings presented in the *Report* and particularly its no-school-effect finding.

The authors consistently entered family background variables first and educational inputs (school factors) last. The result of this procedure, which is referred to in the *Report* as "controlling for background factors," is to assign both the unique and disputed portions of the explained variance to background factors. The decision to enter background factors first (control for) is so critical that we feel it is necessary to consider the rationale behind this procedure. The authors state:

> Since the student's background is clearly prior to, and independent of, any influence from school factors, these background factors can and should be held constant in studying the effects of school variables. Thus, the variation in achievement and attitudes to be explained by school variables is left after variation explained by family background differences is taken out.[27]

We strongly disagree with this statement if independence of school and background factors is interpreted to mean that the spatial distribution or location of families by socioeconomic group is unrelated to the distribution of school facilities by quality. There is abundant empirical evidence to the contrary and any model of residential location that includes the provision of public services also argues otherwise. Moreover, we fail to understand the relevance of the statement if independence is intended to indicate that present school factors cannot *cause* present family background factors.

Hypothesized causal patterns among explanatory variables (at least in the models used by the *Report*) have no bearing on how the variance explained by interaction terms should be allocated. The "prior to" terminology used by the authors appears to be a temporal justification for assigning all of the interaction terms to the family background vector.

The underlying issue is how the jointly explained variance should be partitioned among explanatory variables. Despite the authors' attempts to suggest otherwise, there is no correct way of partitioning it among background and school factors within the context of the study because the samples simply do not contain the information required. These questions can only be resolved by obtaining samples with less correlated input vectors. In these cross-section samples both prior and current influences of family background are intermingled. No temporal interpretation of either the unique or jointly explained variance can be defended. Part of the disputed joint variance may be due to prior background influences, but these issues cannot be resolved within this framework. This difficulty is one major reason several critics of the *Report* have emphasized the importance of obtaining longitudinal data.[28] These questions of prior effects cannot be resolved by fiat.

♦ *The Effect of Ordering—An Illustration*

To illustrate the critical importance of the *Report*'s treatment of interaction terms, we have performed a reanalysis of the data used in the *Report* that displays the effects of reordering the input vectors. However, we wish to make it clear that this is done for illustrative purposes only. For a number of reasons, which we discuss later, we do not believe that these results are a meaningful way of looking at the educational production process.

Our reanalysis relies upon the same samples used in the *Report*'s analysis as reported in the published correlation matrices.[29] We analyzed several different samples, but only a representative case will be presented here— Negro twelfth graders in the North. Findings for the other samples that we analyzed were less dramatic but qualitatively very similar. We attempted to analyze a "median" or "composite" model constructed so as to be independent of any particular one of the many model formulations found in the *Report*. The principal innovation of this model is the inclusion of both teacher and facility measures in the vector of school inputs. The *Report* handles our school vector in what we regard as a peculiar and incorrect manner. In brief, it analyzes the effects of teachers and facilities separately, i.e., it never considers the combined effect of both kinds of school variables in the same model. This practice leads to the use of some rather odd terminology that may account for much of the confusion surrounding the *Report*'s findings about the effect of schools on achievement. In general, the authors of the *Report* do not appear to consider teaching personnel as a school input (al-

though roughly 70 percent of current educational expenditures are accounted for by teachers' salaries).[30] The individual input vectors generally contain more elements than those found in the *Report*.[31]

The amount of variance in individual verbal achievement explained by the vectors of inputs under different ordering schemes is shown in Table 1. The results of this procedure are striking. By adding the unique contributions found in the last column, it is easily seen that only slightly more than half (.0868) of the total explained variance (R^2) is uniquely accounted for by the separate input vectors. In every instance a considerable portion of the explained variance allotted to the first vector entered is actually jointly explained variance. (This is seen by comparing the first and last columns for each input vector.) In all of the samples we reanalyzed in this way, a substantial fraction of the variance assigned by the *Report* to background

TABLE 1

Proportion of Variance Explained By Educational Input Vectors Under Different Orderings—Verbal Achievement of Negro 12th Graders in the North
$$R^2 = .1409$$

| Vector | 1 | Order of Entry | | | 3 |
		2 After S	2 After P	2 After F	
S (school)	.0808	—	.0222	.0560	.0312
P (student body)	.0703	.0117	—	.0420	.0072
F (family)	.0777	.0529	.0494	—	.0484

Components of Input Vectors

1. S = School inputs
   Teachers' ses level
   Teaching experience
   Quality of college attended by teachers
   Degree received
   Teacher preference for middle-class students
   Teacher verbal ability score
   Teacher salary
   Science lab facilities
   Extracurricular activities
   Comprehensiveness of curriculum
   Student transfers in and out
   Attendance
   School location
   Accelerated curriculum
   Promotion of slow learners
   Per-pupil instruction expenditure

2. P = Peer effects
   Proportion white
   Proportion own encyclopedia
   Average hours homework
   Proportion read over 16 books
   Foreign language courses
   Math courses
   Proportion teachers expect best
   Average time guidance

3. F = Family background
   Reading material in home
   Items in home
   Siblings
   Structural integrity of home
   Foreign language in home
   Preschool
   Parents' education
   Urbanism of background
   Homework

factors was actually jointly explained by variables incorporated into our "composite" school vector. Though we do not attach much significance to the finding, our school vector, when entered first, explained more variance than the family vector when entered first. In fact, by all the criteria used in the *Report*, our "composite" school vector appears to be "extremely strong." [32] If the *Report*'s analysis of variance criteria is used, school factors swamp student body factors.

We do not wish to belabor a fairly simple point. It is simply not possible to determine from the sample data how much of the jointly explained variance shown in Table 1 or contained in similar samples is attributable to each of the vectors. It is conceivable, as the *Report* asserts, that all of it is attributable to background (family and student body) factors. It is just as possible that all of it is attributable to school inputs.

◆ A Further Lesson from Reanalysis

Equality of Educational Opportunity provides several partial pictures of education and presents distinctly different models when it examines each aspect of the educational process; for example, the model for teacher effects (page 319) and the model for school facilities effects (pages 308–09). To the extent that the various aspects of the educational process are interrelated (good facilities are found with good teachers), this procedure is improper. Instead, the estimation should be based on a single statistical model including all of the factors believed to have an influence on achievement.

Therefore, within the framework of the *Report*'s analysis, we attempted to perform a more general analysis of variance that would incorporate all of the school and teacher inputs along with family and peer variables. We quickly discovered that there was insufficient independent variation in the school factors to allow inversion of the complete matrix; that is, the variables approached having an identity relationship among them. [33] Although we are uncertain about the cause, we hypothesize that it arises from the combination of two factors. First, the nature of much of the data is such that there are relatively few dimensions of significant variation in the measures concerning schools. This results from the insensitivity of the survey instruments and possibly from a further loss of information through index creation.

Second, there are very few degrees of freedom when school factors are considered. There are a maximum of 269 schools attended by Negro twelfth graders in the North sample (the basis of the reanalysis summarized in Table 1). Nearly half, 133, of these contained fewer than 10 percent nonwhites (a more inclusive category than Negro), and an undetermined number of these had no Negro students at all. Moreover, it seems likely that not all of the sampled schools having Negro twelfth graders are included. [34] School variables (school inputs and student-body characteris-

tics) have the same value for all students in the same school; some variables (per-pupil expenditures, number of school days, school location) are the same for the entire school district. Thus, there are relatively few observations for these explanatory variables. Unfortunately, we were unable to ascertain exactly how many separate school systems and schools are included in the sample used in the *Report* because we did not have access to these samples. When these few independent observations are coupled with the insensitivity of most of the explanatory variables used in the analysis, the likelihood of singularity of the moments matrix becomes great. [35] The swiftness with which the matrix goes singular in our generalized analysis of variance procedure causes us to be suspicious about the overall information content in the sample—a sample from which sweeping conclusions about school effects are made.

◆ Initial Endowments—Some Speculation

The most obvious discrepancy between the conceptual and statistical models is the absence of any measure of initial endowments from the latter. The reasons for omitting initial endowment or, conversely, the conceptual desirability of including it are never discussed in the *Report*. [36] Yet most conceptual models of educational achievement for individual students would include such a concept. Its omission from the statistical model may be an important source of bias. [37] If innate ability is independent of the explanatory variables included in the model, it simply will increase the size of the error term—that is, reduce the amount of variance explained by the model. But, if within the sample experience it is correlated positively with any of the explanatory variables, its influences will be represented by these included explanatory variables. We do not claim to know how initial endowment of individual students is related to the explanatory variables included within the model. However, we would note that innate ability is least likely to be correlated with school inputs which, by construction, are measured only for schools and are most likely to be correlated with individual characteristics and family variables which, by construction, relate to individual students.

There is still another way to view the issue of initial endowments. In discounting the effect of schools, the authors of the *Report* point out that within-school variation in achievement is much larger than between-school variation. [38] They contend this finding demonstrates that school inputs cannot be very important and that most of the differences in achievement are due to family (background) influences that, unlike student body and school inputs, can vary within schools. In reaching this conclusion the authors ignore their own admonition that school inputs may vary within schools even if measured school inputs cannot. [39]

Recognition of initial endowments changes the conclusions that can be reached logically from the overall within- and between-schools analysis of

variance. Family background, peer influences, and school inputs interact with the constraints imposed by innate abilities. Within the same environment (family, peers, school inputs), very large differences in achievement will occur as a result of differences in innate ability. The *Report*'s analyses provide considerable support for this view. None of the published models explain more than 30 percent of the total achievement variance. Our reanalysis of the verbal achievement of Negro twelfth graders in the North, which explains only 14 percent of the variance in achievement, is not atypical. If it is claimed that the *Report*'s models are a correct representation of the educational production process and that all variables included in the model are measured without error, the unexplained variance must be attributable to innate ability. This argument would imply that between 70 and 85 percent of the differences in achievement are due to differences in innate ability. We do not set forth this extreme view as a serious proposition. However, it does raise serious questions about the *Report*'s interpretation of the meaning of the within- and between-school differences in achievement. Furthermore, it emphasizes how little we actually know about the determinants of achievement.

♦ Relative Errors of Measurement

Measurement errors always exist in research, and these are a particularly serious problem in social science research. We already have discussed some errors of measurement resulting from the shortcomings of the *OE Survey*. This discussion deals with a more subtle, but possibly more important, kind of measurement error—systematic departures of the variables actually used in the statistical models from those included in the conceptual model. These departures are regarded as errors of measurement of the independent variables. These measurement errors are important because they systematically bias the results toward the no-school-effect finding.

There are two kinds of errors of measurement of the explanatory variables: (1) historical errors of measurement and (2) contemporaneous errors of measurement. Historical errors of measurement refer to how well or how badly the cross-section data account for the intertemporal influences on achievement. Contemporaneous errors refer to imperfect measurement of influences affecting the achievement of individual students at the time of the *OE Survey*. The unifying thread of the subsequent discussion is that both kinds of measurement errors are largest for school inputs and least for individual and family characteristics. Thus, there is systematic bias of the results due to relative measurement errors.

Historical Errors

The conceptual model views education as a process and depicts achievement as being affected by the entire past history of family, student body, and school inputs. However, the data used in the *Report* relate to a particular point in time and are not cumulative. Useful insights can be obtained from analysis of cross-section data of this kind, but the results must be interpreted very carefully and the relative errors of measurement must be carefully and explicitly considered.

Individual and family characteristics are more in the form of stocks and, hence, are subject to less intertemporal variations than are school inputs, which more closely approximate flows. Thus, use of cross-section measurements of contemporaneous school factors clearly tends to underestimate the total effect of educational inputs on achievement. Better measurement of background factors at a point in time elevates the apparent significance of these factors when compared to the more poorly measured school factors. This source of bias is aggravated by the *Report*'s emphasis on the analysis of the later grades. The authors justify this emphasis on the grounds that the family and student body variables are more reliable for older students (depending as they do on self-reporting). However, this argument cuts both ways. It is equally true that school input measurement errors (in terms of viewing educational inputs over time) increase through time. A student's socioeconomic status or, at least, the relevant educational aspects of his socioeconomic status could easily remain the same throughout his years of school; however, it is virtually impossible for him to spend twelve years in the same school.

If we wish to explain current achievement, we ideally should take into account all the school experiences of students. All high-school students have attended more than one school during their lifetime, and there can be significant differences in the characteristics and quality of the feeder schools serving a large comprehensive high school. A sizable proportion of elementary-school students also have attended more than one school. Thus, even a good cross-section description of school inputs may be a poor estimate of the average quality of the schools attended by the students. The choice of the 12th grade, where family and student body variables are most accurately measured and school inputs least accurately measured, accentuates the bias against school inputs.

Contemporaneous Errors

Achievement pertains to a particular student as do the family and individual characteristics. But school inputs are aggregates, or "macro" variables, pertaining to the *school* attended by the individual student. This may be considered a measurement error for school variables when individual or "micro" relationships are analyzed. For example, well-equipped science labs may be of little value to students enrolled in a business course and extracurricular activities may provide few benefits to a student from a low-income family who must work after school. Aggregation of school inputs reduces the apparent explanatory power of such school factors (as compared to the

case where school factors are measured without error). This kind of measurement error, which could be called school-input heterogeneity, is most serious for comprehensive high schools. The curricula of many of these are so segregated that they might better be thought of as separate educational institutions, except when it comes to fielding athletic teams. Tracking, within both comprehensive high schools and elementary schools, is still another cause of school-input heterogeneity. In short, students attending the same school may receive markedly different educational inputs.

♦ Peer Effects and School Quality

Given the probable magnitude of historical and contemporaneous errors in the measurement of school inputs, it is essential to consider the relationship between the average characteristics of students attending a particular school (peer effects) and the quality of that school. As Table 1 indicates, the school input and student-body (peer effects) vectors are highly intercorrelated. The school vector entered first explains 8.08 percent of the total variance. Similarly, the student-body vector entered first explains 7.03 percent. However, the incremental contribution of the student-body vector (the school input vector entered first) is only 1.17 percent and the incremental contribution of the school input vector (the student-body vector entered first) is only 2.22 percent. Between them (entered before the family background vector) they explain 9.25 percent of total variance. Of this total 5.86 percent is jointly explained by the two vectors. Only 3.39 percent (2.22 plus 1.17) is uniquely assignable to each of the input vectors.

It is not hard to justify the view that families who are more educationally concerned or who can afford to spend more on the education of their children may systematically choose to send their children to the best schools. Thus, the best schools, in terms of both broadly and narrowly conceived educational inputs, probably will have disproportionate numbers of students with well-to-do and educationally concerned parents. Under these circumstances, student body characteristics will be highly correlated with real differences in school inputs. If school inputs are measured poorly, it is possible that these student body characteristics may be the best measure of the quality and quantity of school inputs. This assertion cannot be assessed in any adequate way using the *OE Survey* data, but if it is true, part of the effect that the *Report* attributes to peer or student-body effects may be indirect measurements of the effect of school inputs.

The above argument leads to a simple conclusion: some part of the disputed variance between student body and school inputs is undoubtedly attributable to school inputs. Indeed, some part of the uniquely explained variance in Table 1 attributed to family and student body variables actually may be attributable to school inputs. Finally, it should be noted that better specification of the school vector (inclusion of both facilities and teachers at

the same time) as in Table 1 considerably reduces the explanatory power (unique contribution) of student-body factors that are deemed "important" in the *Report*.

♦ Twelfth Grade?

The *Report*'s emphasis on later grades affects its analysis through more than time-related errors of measurement. First, it is likely that the independent effects of schools are strongest and most easily identified in the earlier years of schooling. Second, modeling an elementary school is a more tractable problem than modeling a comprehensive junior high or high school. It would be extremely difficult to describe exhaustively the school inputs at the high-school level, even if ideal measures of school inputs relevant to each individual student were available. No one could describe the *OE Survey* data on school inputs as ideal. These problems compound the bias against school factors.

The *Report*'s authors were obviously most interested in the 12th grade; however, they claim to have tested relationships in other grades and found that school factors had little effect on the result. At one point they conclude:

> At grades 3 and 1, little variance is accounted for either by school characteristics or student body characteristics. This result, in which no variables account for much of the variance in achievement, is true throughout the analysis for grades 3 and 1, despite the large school-to-school variations shown in Tables 3.22.1 and 3.22.2.[40]

What the authors fail to point out in this discussion (except in a footnote to Table 3.22.1) is that almost no school or student-body variables are analyzed for the earlier grades. Of the eleven school characteristics included for the 9th and 12th grades, only per-pupil expenditure (which pertains to the entire school district and is strongly affected by whether the district is simply an elementary-school district or a unified one), volumes per student in the library, school enrollment, and location (city, suburb, town, county) are included in the 6th, 3rd, and 1st grade analysis. Of the six student-body characteristics, only the proportion whose families own encyclopedias, an index of student transfers, attendance rates, and teachers' perception of student body-quality were included. We were not too surprised with the result when we discovered the limited number and scope of variables used in the analysis for earlier grades.

The selective nature of the school population (dropouts) is another difficulty associated with the *Report*'s emphasis on the 12th grade. In 1960 only 82 percent of whites aged sixteen and seventeen and 73 percent of Negroes in the same age group were enrolled in schools.[41] These figures indicate a systematic difference in nonenrollment by race. Moreover, nonen-

rollment almost certainly is correlated with factors such as ability, achievement, socioeconomic characteristics, and residence. Little is gained and many problems are introduced by concentrating on grades where this problem is most severe.

◆ *Some New Evidence*

Still another kind of empirical evidence is used in the *Report* to support its no-school-effect finding. The authors contend that, if schools did have an effect, the relative variance between schools would increase over time. Since the between-school variance remains fairly constant among grades, they conclude that schools must have little effect on student achievement. (Note that direct comparisons are only valid for grades one, three, and six where school size is roughly constant.)[42] This interpretation of the finding of approximately constant between-school variance over the years of schooling assumes the distribution of output among schools is stationary. That is, it implicitly assumes that schools with low mean achievement in the 1st grade continue to have low mean achievement in later grades; and vice versa for schools with high mean achievement. Information obtained from subsequent analysis of the *OE Survey* data is relevant to this conclusion.[43]

Analysis of mean achievement scores for a sample of schools in the Northeast and the Great Lakes region indicated that mean 1st-grade and mean 6th-grade verbal test scores were not highly correlated. For 100 schools that contained more than four Negroes, the simple correlation of the mean Negro 1st-grade verbal score with the mean Negro 6th-grade verbal score was .29. For 198 schools with over four white sixth graders, the correlation for whites was .36. This hardly supports the view that the development path is unaffected by school inputs when one considers that neighborhood schools tend to insure similarity of social and economic characteristics of first and sixth graders.

One possible explanation for this phenomenon is that the 1st-grade test is not a good test. This, among other things, would imply that the intergrade comparisons of variance are not meaningful. An alternative explanation is that schools do have an effect on students, and, while the total variance tends to be similar, the position of a given school within the distribution is altered by the inputs of the school to individual students' education. Again, the intergrade comparisons of total variance are not meaningful. Conclusions requiring time series data that are made from cross-section data are unwarranted. The *Report*'s inferences based on intergrade comparison of between-school variance are questionable.

◆ *Model Specification—The Implications of Linearity*

There are a number of issues relating to the model's overall specification and, in particular, to the choice of functional form. The linear, additive specification used by the authors has serious limitations. First, there is a dimensionality problem. For example, the effects of guidance counselors surely must be related to the number of students. Yet the *Report* introduces these variables in their original form. In an additive model these dimensionality problems are not accounted for by the addition of school size. Second, there are possibilities of scale problems in the economist's sense. In particular, it has been argued that scale economies are likely to exist in high schools.[44] If important economies of scale do exist, the simple linear form is incorrect. Third, it does not allow for the interaction between inputs. The effect of a given input is the same whether or not any other inputs are absent or are found in such abundance as to be superfluous. Fourth, the linear form implies that the marginal effect of a given input is the same regardless of the level of usage. Adding one guidance counselor has the same impact on student achievement when the change is from 0 to 1 as when it is from 500 to 501. Certainly few people would hold to this implicit assumption of a constant marginal product of inputs.

The principal justification (rationalization) of the linear form is that many mathematical functions look linear over a small range and all of the *Report*'s findings must be interpreted within the rather limited range of the explanatory variables. Moreover, since statistical estimation techniques are most highly developed for linear models, a linear function is generally chosen in the absence of strong a priori views favoring alternative specifications. Nonetheless, it is important to understand fully the strong implications of the functional form selected by the *Report*'s authors and its possible limitations.

◆ *Analysis of Variance and Public Policy*

There is one final consideration, the appropriateness of analysis of variance for studying the range of policy questions undertaken. For policy purposes, it is desirable to identify and evaluate the impact of potential policy instruments on achievement. Evaluation of alternative policies necessarily involves consideration of both the effects on output of different changes in the inputs to the production process and the costs of these changes. An analysis of variance provides almost no insight into these questions. It does not even give the direction, let alone the magnitude, of the effect that can be expected from a change in inputs. What are needed for policy purposes are estimates of the parameters of the statistical model, i.e., the a_i's in Equation 2. There is a considerable difference between the concepts and methods

employed in partitioning the variance among sets of variables and those employed in estimating slope coefficients or elasticities for individual variables.[15] The latter estimates are useful in identifying potential policy, the former are of considerably less value to the policy-maker.

The closest the *Report* comes to identifying policy instruments are its estimates of the "unique" contribution of individual variables to explained variance. This procedure amounts to carrying out the specialized analysis of variance for individual variables, rather than for vectors. (The *Report* defines unique contribution when the given variable is the last one added to the regression equation.) As before, the outcome of this procedure is dependent upon the amount of intercorrelation present in the system. If high intercorrelations exist, the concept of "unique contribution" is not of much help in identifying policy instruments.[16] Instead, such an analysis is largely an indication of which variables are most orthogonal (are least highly correlated with other explanatory variables). By contrast, multiple regression analysis focuses on estimates of the independent effects of different explanatory variables. Multicollinearity tends to reduce the precision of the parameter estimates (increase the standard errors), but least squares techniques prove quite robust even when the explanatory variables are intercorrelated. The regression coefficients have remained unbiased in the presence of intercorrelations among input vectors or individual variables.[17] However, when there is a significant degree of multicollinearity present, there is no simple relationship between the regression coefficient and the amount of variance explained by the variable. Even if all explanatory variables were truly independent (not correlated with any other explanatory variable), the analysis of variance format would not be the most useful mode of analysis. The proportion of explained variance does not identify policy instruments and gives little indication of the extent of policy leverage provided by different variables. Parameter estimates are much more useful in this respect. The really interesting questions involve the effects of changes in inputs to the educational process. Explained variance, whether an orthogonal component or not, is simply not a very interesting concept either to the policy-maker or the statistician.

SUMMARY AND CONCLUSIONS. *Equality of Educational Opportunity* has not served us well as a policy document. Distracted by the allure of basic research into the educational production process, the Office of Education failed to provide an authoritative response to the Congressional request for data on educational opportunities. The extent to which minority groups are systematically discriminated against in the provision of educational inputs is still unknown. This is a serious matter since the correction of input inequalities is a logical and necessary first step in insuring equality of opportunity for minorities.

The *Report*'s failure to provide a definitive answer to the Congres-

sional question concerning inequality in the provision of educational inputs may be its most fundamental weakness as a policy document. Yet our evaluation of the *Report* is concerned primarily with serious flaws in its basic research into the educational production process. The reason for this is that the "findings" of the *Report*'s basic research have been widely acclaimed as its principal policy contribution. We contend that the *Report*'s analysis does not provide reasonable tests of the hypotheses attributed to it. In fact, the "findings" could have been the result of the analytic methods combined with systematic errors of measurement rather than any underlying behavioral relationship.

The large and systematic differences between the *Report*'s implied conceptual model and its actual statistical models of the educational process are its most damaging analytic shortcomings. Thus, while serious interpretative problems are raised by the specialized analysis of variance procedure used in the *Report*, its principal failing is found in its inadequate statistical models of the educational process. This inadequacy arises in large part from using poor or incomplete data that bias the empirical results toward overstating the effect of family background and student-body inputs and understating the effect of school inputs.

Although education must be viewed as a cumulative process, the *OE Survey* data contain no historical information. This weighs most heavily on school factors because contemporaneous school data are less adequate surrogates of historical influences than are contemporaneous family background measures. The emphasis on later grades exaggerates the bias toward a no-school-effect finding.

Errors also exist in the measurements of contemporaneous inputs to the production process and these are greatest for school inputs. The most critical error is the failure to collect and use data on school inputs for individuals. This omission is especially serious in later grades where the students can choose different curricula and are more completely separated. Family inputs, which are measured for the individual, necessarily have less error in a model of individual achievement and, consequently, their apparent explanatory power is raised in comparison to the more poorly measured school factors.

The neglect of innate ability in the statistical model may also bias the estimated influence of family factors upward. If innate ability is partially hereditary and social mobility exists through ability, this excluded input to the production process will be correlated with family measures, and the statistical results will be biased toward showing stronger family effects than actually exist. In any event, the *Report*'s use of differences in within- and between-school variance in achievement as prima facie evidence for the unimportance of schools in determining achievement is clearly unreasonable when the potential role of initial endowment is recognized.

In this paper we have said very little about what educational produc-

tion functions actually look like. Instead our discussion is limited to a demonstration that the *Report* does not answer this question. More attention is given this question in an empirical analysis of the education process by one of the authors.[48] This analysis based on *OE Survey* data is limited by many of the data problems which hampered the *Report*'s authors. Even so, the empirical findings of that study indicate that alternative, and we contend more appropriate, models and methodology lead to important differences in empirical results. In particular, these alternative models tend to confirm the view advanced in this paper that the *Report*'s "no school effect" is due in considerable degree to its method of analysis.

Much of the preceding discussion is essentially negative. It recommends what policy-makers should not do: they should not rely very heavily on *Equality of Educational Opportunity* and in particular on its analysis of the relationship between inputs and outputs in designing educational policy. As a pioneering piece of social science research, the *Report* deserves considerable praise. However, as a policy document, it must be evaluated differently. In this guise it is potentially dangerous and destructive.

If the *Report*'s analysis of the educational process cannot be believed, what is the policy-maker to do? The one incontrovertible finding from the *OE Survey* is that the median educational attainment of blacks is considerably below that of whites. The average 12th-grade Negro in the North (who is still in school) is achieving at the 9th-grade level of his white counterpart.[49] This divergence increases when other regions are considered, reaching an apogee in the rural South where the achievement of 12th-grade Negroes lags five years behind that of Northern white twelfth graders. The existence of such sizable differentials is amply demonstrated by the *Report;* the best ways to eliminate the differentials are not.

Since we do not believe adequate knowledge for program design exists, we will not even speculate on the best mix of educational resources or programs. The authors of the *Report* have performed a valuable contribution in again reminding educational policy-makers that the production of educational output does not stop at the school door. An effective program for increasing the educational achievement of culturally deprived children, be they white or black, would almost certainly require a mix of school and nonschool programs. If publication of the *Report* has made educational policy-makers think about education in this broader framework, it will have provided an invaluable service. However, if these policy-makers conclude the *Report* provides an adequate basis for choosing among alternative programs and mixes of expenditures, it will have done a grave disservice. It simply does not provide satisfactory answers of the kind widely ascribed to it.

If, as we contend, the information does not presently exist for designing optimal or even efficient educational programs, what can be done? An admission of ignorance should not be interpreted as a plea for inaction.

Nothing would be more disastrous. The problems of low educational achievement among the culturally deprived, and particularly blacks, documented by the *Report* and other studies, will not disappear simply because the knowledge needed to design "efficient" programs is unavailable. The magnitude of these problems is so great as to demand immediate and large-scale action, even if many efforts will prove to be "inefficient."

What is demanded by a situation such as this is a process of radical experimentation, evaluation, and research. Additional large-scale research programs, similar in scope to that reported in *Equality of Educational Opportunity*, would be of considerable value in identifying promising areas for program development, experimentation, and evaluation. This is particularly true insofar as subsequent studies are able to overcome some of the most serious deficiencies of the *OE Survey*. The most prominent of these are the primitive measurements of school inputs and lack of longitudinal data.

It is beyond the scope of this paper to suggest the precise nature of the experimental activity. Indeed, it is beyond our resources and competence. However, we can identify several desirable characteristics for such a program. It is essential that the experimentation be truly radical: that is, involve a wide variety of educational practices and explore ranges of input variation in both novel and traditional educational techniques presently not found in the public schools. As is noted in the preceding discussion, the range of educational practice examined by the authors of the *Report* is exceedingly limited. This is not because the authors wanted it this way, but rather because that is the way it is. For example, the differences in student-teacher ratios examined by the *Report* were very small. Moreover, larger classes more often characterized situations where the benefits of small classes would seem greatest. We doubt that the analysis included any situations where culturally deprived students had been saturated with educational inputs over their educational lifetimes. Yet without such experience the effects of providing much higher levels of educational expenditures cannot be evaluated. It seems likely that very small changes in the level of educational inputs may have little effect, but that there may be a threshold where increased resources become effective. Similarly, the experimental program should be large-scale. Many different kinds and combinations of programs should be tried. Finally, experimental programs must last long enough to operate on the entire process of education. Massive inputs of educational resources in junior high or high school may be quite ineffective if the history has been one of neglect.

Experimentation in nonschool programs designed to increase educational achievement may be more difficult to carry out. Still, a program of large-scale experimentation such as we envisage ideally should attempt to include such programs. At minimum, the school should be construed very widely and every effort should be made to experiment with programs that are intended to extend their influence into the community. These would in-

clude preschool activities, day care, recreational and educational summer programs, after-school adult education activities, and programs designed to involve the parents to the greatest possible extent.

Evaluation is the final and most critical aspect of any well-designed program of experimentation. Without adequate evaluation, it will not be possible to determine what mix of experimental programs worked and why. Education is a highly complex process and adequate evaluation would be both expensive and difficult to design and carry out. Still, the potential benefits from high-quality evaluation, through increasing the efficiency of ongoing programs, are very great.

Notes

1. The authors would like to acknowledge the large number of constructive criticisms and suggestions on both the content and style of earlier drafts of this paper by persons too numerous to mention here. A special note of thanks is due Joseph J. Persky, John Jackson, Thomas F. Pettigrew, David Cohen, Leonard Rapping, Molly Mayo, and Frederick Mosteller, all of whom made particularly helpful suggestions. Finally we would like to express our appreciation to James S. Coleman, who was kind enough to bring to our attention several errors of interpretation of the *Report*'s analyses and findings contained in an earlier draft. Of course, any errors that remain are the sole responsibility of the authors.

2. James S. Coleman, Ernest Q. Campbell, Carol J. Hobson, James McPartland, Alexander M. Mood, Frederic D. Weinfield, and Robert L. York, *Equality of Educational Opportunity.* Washington, D.C.: U.S. Government Printing Office, 1966. This is subsequently referred to as the *Report.*

3. While there might be some doubt among those who have digested the *Report*'s tables and analysis as to whether the no-school-effect conclusion is the principal finding, it is clear that the majority of commentators on the *Report* have reached this conclusion. Moreover, the summary chapter, the major source of discussions about the *Report*, states:

> The first finding is that schools are remarkably similar in the way they relate to the achievement of their pupils when socioeconomic background of the students is taken into account. It is known that socioeconomic factors bear a strong relation to academic achievement. When these factors are statistically controlled, however, it appears that differences between schools account for only a small fraction of differences in pupil achievement. [*Report*, p. 21.]

This view of the *Report*'s findings has found its way into the professional journals as evidenced by the Editors of the *Harvard Educational Review* who state:

> . . . Coleman's analysis of the survey data suggests that the traditional remedies proposed by educators—increased expenditures, reduced class size, improved facilities, ability tracking—will make little dent, for these factors evidently exercise almost no independent effect on pupil achievement when family background variables are controlled. [*Harvard Educational Review*, Vol. 38, No. 1, Winter 1968, p. 85.]

The final demonstration of the pervasiveness of the no-school-effect finding is the fact that it was promulgated at Senate hearings by Daniel P. Moynihan. He noted that Dr. Coleman

> . . . found that the quality of schools could not explain differences in achievement, excepting of a relatively low order. For example, for Negro students in

the urban North, expenditures per pupil on instruction could only account for two-hundredths of 1 percent in the variation of achievement of ninth grade students. This is obviously nothing significant. [U.S. Senate, *Federal Role in Urban Affairs: Hearings before the Subcommittee on Executive Reorganization of the Committee on Government Operations*, Part 13, December 13, 1966, p. 2692.]

4. Although it is difficult to categorize and document the *Report*'s findings, two additional views stand out. First, the *Report* is used to support the contention that schools attended by whites and minority students are not very different. This position is typified by Daniel P. Moynihan's statement that "despite our expectations, by and large the quality of school facilities available to minority children in this country are not significantly different from those available to the majority." [*Hearings before the Subcommittee on Executive Reorganization of the Committee on Executive Reorganization of the Committee on Government Operations*, 89th Congress, Second Session, December 30, 1966. Appendix to Part I, p. 2692.] This finding, we believe, is more a product of data problems (discussed in Section I, below) than of the actual distribution of facilities.

   Second, close to the no-school-effect finding in terms of frequency of citation is the integration finding, i.e., integration is good because Negroes learn more in integrated schools. An example of this interpretation of the *Report* is found in Irwin Katz's letter to *Science*, May 12, 1967, p. 732. Many of the methodological difficulties discussed in this paper which seriously undermine the *Report*'s no-school-effect finding apply with equal force to the integration finding. A more detailed discussion of the integration finding, particularly as it relates to the subsequent report by the U.S. Commission on Civil Rights, *Racial Isolation in the Public Schools*, Vol. 1 (Washington, D.C.: U.S. Government Printing Office, 1967), may be found in Eric A. Hanushek, "The Education of Negroes and Whites," unpublished Ph.D. dissertation, Department of Economics, Massachusetts Institute of Technology, August, 1968.

5. Samuel Bowles and Henry Levin raise many similar issues in "The Determinants of Scholastic Achievement—An Appraisal of Some Recent Evidence." *Journal of Human Resources*, Vol. III, No. 1, Winter 1968.

6. James S. Coleman reaches the same conclusion when explaining the broader construction of the mandate developed by the Office of Education. In a recent article he states:

> The Congressional intent in this section is somewhat unclear. But if, as is probable, the survey was initially intended as a means of finding areas of continued intentional discrimination, the intent later becomes less punitive-oriented and more future-oriented; i.e. to provide a basis for public policy, at the local, state, and national levels, which might overcome inequalities of educational opportunity.

James S. Coleman, "Equal Schools or Equal Students?" *The Public Interest*, No. 4, Summer 1966, p. 70.

7. James Coleman, writing eloquently on the concept of educational opportunity, identifies five kinds of inequality. These include the inequality inherent in racial segregation and that relating to a variety of output measures. James Coleman, "The Concept of Equality of Educational Opportunity." *Harvard Educational Review*, Vol. 38, No. 1, Winter 1968.

8. The expenditure measure has the advantage of providing a weighting scheme for real resources, i.e., resources are weighted by their costs. This allows an easy comparison of input quantities among schools. One of the foremost problems with this technique, however, is just this weighting scheme because, if large price differentials exist, direct comparisons are of limited value. Indeed, large input price differences are known to exist among regions, e.g., by teacher salary differences. This problem can be alleviated by either restricting comparisons to a uniform price area, such as a city, or constructing price indices. On the other hand, while real input measures avoid the problem of differences in prices, there is no

natural set of weights which can be used to create a single real resource measure. Thus, comparisons of differences in real inputs ideally require far more knowledge about the effect of these inputs on output than now exists and considerable judgment in interpreting the relative importance of various factors. (The ideal weighting scheme would account for the value of an input in producing education. If schools are operated efficiently, the expenditures on various inputs will provide such a weighting scheme.)

9. There are several ways in which these data on outputs could be obtained. First, one could sample the schools attended by various minority groups and obtain average achievement levels by schools. Alternatively, one could break away from the reliance on schools as the sampling unit and sample the population, as in the *Current Population Survey* of the Census Bureau. Finally, since many school systems conduct extensive testing programs, one could collect standardized achievement test scores already available in the schools. This would provide considerable information at a small marginal cost, especially if undertaken in conjunction with a survey of inputs. The two most populous states (California and New York) currently conduct statewide testing programs. In fact, on p. 105, the *Report* states that only five percent of the Negro and one percent of the white elementary students in the United States attend schools that do not give standardized achievement tests. Over ninety percent are tested two or more times. Similarly, less than five percent of the secondary students are not tested.

10. Chapter 2 of the *Report* contains many tables describing the schools attended by whites and the several minority groups considered by the *OE Survey*. These data have been interpreted widely as indicating the absence of all but very small differences in the quality of schools attended by whites and Negroes, at least within the same region. However, the combination of considerable fragmentary evidence to the contrary and our serious misgivings about the representativeness of the *OE Survey* and the quality of the data obtained on school inputs causes us to question this "finding." A number of expenditure studies have documented very large suburban-central city differences in per-pupil expenditures. Increasingly, central city-suburban is a euphemism for black-white. For evidence on these expenditure differences see: Dick Netzer, *Economics of the Property Tax* (Washington, D.C.: The Brookings Institution, 1966), Chapter 5; and U.S. Commission on Civil Rights, *Racial Isolation in the Public Schools* (Washington, D.C.: U.S. Government Printing Office, 1967), pp. 25–31. Several studies have produced evidence of significant expenditure inequalities by racial groups within the same school system: Martin T. Katzman, "Distribution and Production in a Big City Elementary School System," unpublished Ph.D. dissertation, Yale University, 1967; The Education Committee, Coordinating Council of Community Organizations, *Handbook of Chicago School Segregation, 1963* (Chicago: The Committee, August, 1963); Patricia Cayo Sexton, "City Schools," *The Annals of the American Academy of Political and Social Science*, 352 (March, 1964), pp. 95–106; California, Governor's Commission on the Los Angeles Riots, *Violence in the City—An End or a Beginning?* (Los Angeles, California, December 2, 1965), pp. 49–61.

11. The Southern states are: Alabama, Arizona, Arkansas, Florida, Georgia, Kentucky, Louisiana, Mississippi, New Mexico, North Carolina, Oklahoma, South Carolina, Tennessee, Texas, Virginia, and West Virginia.

12. Hanushek, "The Education of Negroes and Whites."

13. The *Report* handled internal nonresponses in two ways. If the question was to be used in an index, the sample mean was assumed to be the correct answer. If the question was to be used as a separate variable, the covariance matrix for the observations was calculated on the basis of complete observations, and individual elements were weighted in such a manner as to arrive at population figures. Cf. *The Report*, p. 572. Both corrections yield considerable errors in the face of systematic nonresponse.

14. Analysis and editing of the *OE Survey* data provide many obvious cases where teachers, principals, and superintendents coded their answers to the left rather than the right as required of such numerical responses (left justification instead of

right justification). Machine processing of these incorrectly coded answers creates a decimal point error that can lead, if not corrected, to sizable errors in any statistical analysis. The frequency of such decimal errors on student and/or teacher population questions answered by the principals in the above 300 school sample was over 10 percent. Decimal errors of this kind often can be corrected by cross-checking with other size indicators.

15. A preliminary check of the data made by comparing question responses in schools with two principals does provide a simple reliability test for different variables. As one would expect, faulty response, though not as serious on the multiple-choice answers, is an ever-present problem. The *Report* mentions one very small reliability test that the Educational Testing Service carried out. However, this seems quite inadequate.

16. For example, in terms of problems with nonschool data, each student completed a multiple-choice question concerning father's occupation that included one large category mixing traditionally blue- and white-collar occupations.

17. The basic notion of a model pertains to a simplification of reality that allows us to analyze particular aspects of a process—in this case the production of education. The conceptual model represents the theoretical backdrop for the empirical section which is the statistical testing of the hypothesized relationship. Though more will be said about the statistical aspects of the model, proper testing of a model calls for specification of all of the major influences on the dependent variable (achievement). The fact, however, that a model is a simplification of actual conditions implies that we will not be able to explain or predict perfectly the level of achievement. Statistical theory provides us with criteria for selecting models and judging their merits.

18. Proposals of the latter kind are suggested by James Coleman in an article published in the *Public Interest*. There he argues in favor of replacing the family environment of the disadvantaged child as much as possible "with an educational environment—by starting school at an earlier age, and by having a school which begins very early in the day and ends very late." Coleman, "Equal Schools or Equal Students?" p. 74.

19. The *Report* comes close to describing a very similar conceptual model on page 295. However, since the authors never refer to it in subsequent sections, it is difficult to believe that they actually attach much meaning to it.

20. The true (population) values of the a_j's are hypothesized to be the same for every individual in the population considered.

21. Variance is the average squared deviation of individual observations from the sample mean, i.e., variance $(A) = \Sigma[(A_i - A)^2]/N$ where A is the sample average of the N observations of A_i. Therefore, it is the standard deviation squared and, in a loose way, it measures the dispersion of individuals from the mean performance level.

22. Least squares regression technique is a method of estimating the parameters, a_j, of Equation 2. The basic criterion in estimation is the minimization of the sum of e_j^2. If certain conditions about properties of the e_j hold (e.g., that the e_j's are uncorrelated with the X_{ij}'s), the technique of least squares is shown to possess some desirable attributes pertaining to the estimates of the a_j's. Additionally, it is possible to relate the size of the residuals and the variance of the dependent variable in a manner that gives some feel for "how good" the model is. This measure, R^2, or the squared multiple correlation coefficient, must assume values between zero (no explained variance) and one (all of the variance explained).

23. Explanatory variables frequently are referred to as independent variables. It is important to understand that in social science research these so-called independent variables are seldom truly independent in the statistical sense of being uncorrelated with one another.

24. Relationships between the average income or social class of the school and school inputs or per-student expenditures have been found by Patricia C. Sexton, *Education and Income* (New York: Viking Press, 1961), and Martin T. Katzman, "Distribution and Production in a Big City Elementary School System."

25. The general statistical term for correlations among explanatory variables is multi-

collinearity. In the following discussions we are concerned with a specific type of multicollinearity, i.e., intercorrelations among the input vectors of Equation 1 rather than intercorrelations among the specific variables representing each vector.

26. The explained variance can be decomposed into a part that is uniquely explained by individual vectors or variables and one that is jointly explained by more than one vector or variable. This jointly explained component, which can also be decomposed into specific combinations of inputs, is referred to subsequently as an interaction term.

27. The *Report*, p. 330.

28. For example, see Bowles and Levin, "The Determinants of Scholastic Achievement," or Christopher Jencks, "Education: The Racial Gap," *The New Republic* (October 1, 1966), p. 21.

29. Supplemental Appendix to the *Survey on Equality of Educational Opportunity.* Washington, D.C.: U.S. Government Printing Office, 1966.

30. For example, James Coleman in his article published in the *Public Interest* summarizes the findings of the study as follows:

> Even the school-to-school variation in achievement, though relatively small, is itself almost wholly due to the social environment provided by the school, the educational backgrounds and aspirations of other students in the school, and the educational backgrounds and attainments of the teachers in the school. *Per pupil expenditure, books in the library, and a host of other facilities and curricular measures show virtually no relation to achievement if the "social" environment of the school—the educational backgrounds of other students and teachers—is held constant.* [(Coleman's emphasis) Coleman, "Equal Schools or Equal Students?" p. 73.]

31. Our procedure was to include the maximum number of explanatory variables in each vector. Because of the multicollinearity among explanatory variables of each vector and the limited information contained in the sample generally, substantially fewer than all independent variables could be included.

32. This finding does not appear in the *Report* because of the separate analysis of facilities and teachers.

33. Our analysis considered white and Negro students in the North for grades six and twelve.

34. The authors did not use all of the survey data in their analysis of variance procedure. Rather, they randomly sampled 1,000 students from each of five strata included in the North sample (nonmetropolitan North, nonmetropolitan West, metropolitan Northeast, metropolitan Midwest, and metropolitan West). The probability of a school's being selected was dependent on its 12th-grade Negro enrollment; consequently the North sample was weighted heavily toward schools with predominantly Negro enrollment. It is difficult to understand why these sampling methods were used since presently available computers are unawed by large sample sizes. Whatever the reasons for the internal sampling, the procedure amplifies our earlier observations about the desirability of choosing a single research strategy and concentrating efforts on collecting the information and tailoring the analysis to answer a particular set of questions.

35. The least squares regression technique requires inverting the moments matrix of the data. (This is close to the simple correlation matrix of the variables.) Inversion is impossible if a linear identity exists among the rows or columns of the matrix. However, as a practical matter, the procedure is stopped at some point before perfect linear dependency due to the round errors in computation.

36. This is a slight overstatement. The Technical Appendix to Section 3.2 states, "they [the models] do not include differences in native endowments, *which of course must also be considered part of family background, though* an unmeasured part [Our emphasis]." We would question this assertion about the necessary link between native endowment and family background, as they certainly will not be perfectly correlated. However, this is not a serious issue since there is little evidence that the authors of the body of the *Report* acknowledge the possibility of

differences in innate abilities. A plausible hypothesis would be that the sections were written independently by different authors.

37. Bias is used throughout in the formal statistical sense (expected value of estimated coefficient $\hat{a}_i \pm$ true a_i). However, this generally implies erroneous results in the analysis of variance format used in the *Report*.

38. Cf. the *Report*, page 296, and especially Table 3.22.1.

39. The *Report*, page 295. The extreme case of within-school variance that could result from differences in school factors would be found in the comprehensive high school. However, the differences can also exist in elementary schools where there is more than one class for a given grade.

40. The *Report*, p. 304.

41. The *Report*, p. 447.

42. See The *Report*, p. 296.

43. Hanushek, "The Education of Negroes and Whites."

44. Economies of scale exist when a proportional increase in all inputs yields a more than proportional increase in output. For evidence on the existence of scale economies see: John Riew, "Economies of Scale in High School Operation," *The Review of Economics and Statistics* (August, 1966), pp. 280–7.

45. The elasticity of an independent variable is the percentage increase in the dependent variable (achievement) that can be expected from a one-percent increase in the given independent variable. The elasticity of A with respect to X is $(dA/A)/(dX/X)$. In the linear model such as Equation 2, dA/dX equals the regression coefficient for X.

46. This is especially true if, as commented on before, many different models are used to analyze different but related aspects of the educational process.

47. Note that we consider the regression coefficient, a_i, throughout the discussion. This differs from the beta or standardized regression coefficient often used by sociologists: beta equals the regression coefficient times the ratio of standard deviations of the independent and dependent variables. When the explanatory variables are orthogonal, the beta coefficient is directly related to explained variance. With multicollinearity, there is no direct relationship.

48. Hanushek, "The Education of Negroes and Whites." This analysis of the *OE Survey* data relied upon a sample of metropolitan elementary schools in the Northeast and Great Lakes regions. In order to minimize the biases due to incomplete data, the school, rather than the individual, was used as the basic observational unit. Thus, the output of the educational process is defined as the mean achievement level (of sixth graders) within a school and the inputs are aggregate student body and school characteristics. Separate functions were estimated for white sixth graders and for black sixth graders. For both blacks and whites, differences in school quality as measured by average teacher verbal score, average teacher experience, and percent of students with a nonwhite teacher during the previous year exhibit a significant influence on educational attainment.

49. The *Report*, Table 3.121.1.