DOCUMENT RESUME

ED 026 721 24 EA 001 956

By-Coleman, James S.
The Evaluation of Equality of Educational Opportunity.
Johns Hopkins Univ., Baltimore, Md. Center for the Study of Social Organization of Schools.
Spons Agency-Office of Education (DHEW), Washington, D.C. Bureau of Research.
Report No-R-25
Bureau No-BR-6-1610
Pub Date Aug 68
Grant-OEG-2-7-061610-0207
Note-41p.
EDRS Price MF-$0.25 HC-$2.15
Descriptors-Academic Achievement, Analysis of Variance, Educational Equality, Educational Opportunities, Educational Policy, Educational Resources, *Equal Education, *Evaluation Methods, *Federal Programs, Research Design, Resource Allocations, *School Segregation, Statistical Analysis
Identifiers-Civil Rights Act of 1964, *Equality of Educational Opportunity

An author of the 1966 USOE report titled "Equality of Educational Opportunity" discusses some of the criticisms leveled against the study. The most serious charge is that the study attempted to do too much. By analyzing resource inputs and levels of achievement and by trying to analyze the effects of various inputs on achievement, the study failed to do well the minimum requirement--a careful study of inputs that would provide an adequate measure of educational equality. On the other hand, the author argues that the major virtue of the study is that it did not accept the definition of educational equality based on input levels; therefore, its major impact has been to shift policy attention from the traditional focus on comparison of inputs to a focus on output and on the effectiveness of inputs for bringing about changes in putput. The seemingly simple and measurable concept of inputs is further called into question by pointing out the difference between inputs as disbursed and inputs as received by the child. Other criticisms of basic design and specific criticisms of statistical methodology are discussed and met. Although the report is admitted to contain imperfections, it is defended as a major step in the development of policy-related social research. (TT)

THE JOHNS HOPKINS UNIVERSITY

REPORT No. 25

THE CENTER FOR THE STUDY OF SOCIAL ORGANIZATION OF SCHOOLS

THE EVALUATION OF EQUALITY OF EDUCATIONAL OPPORTUNITY

JAMES S. COLEMAN
THE JOHNS HOPKINS UNIVERSITY

AUGUST, 1968

In July of 1966, the U.S. Office of Education issued a report titled "Equality of Educational Opportunity," to fulfill a provision of the Civil Rights Act of 1964 which read as follows:

> The Commissioner 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.

This request was one of the first specific requests made by Congress for social research that might provide a basis for policy. It is a kind of governmental interest in information about the functioning of society that compares to its interest that began some years ago in information about the functioning of the economy. As such, it is likely to increase as national policy becomes increasingly concerned with social institutions, an increase that is already foreshadowed by such developments as the work toward a social report of the President that is now being carried out in the Department of Health, Education, and Welfare.

As a consequence, it becomes important to examine retrospectively this attempt to address social research to social policy, as a way of learning, as social scientists, the problems and pitfalls of such activity, and of learning how best to carry it out. Such activity has not been the central focus of applied social research, and as a consequence, it raises new problems of design and analysis.

I propose, then, to make such a retrospective examination. To do so as one of the authors of the report carries both advantages and disadvantages. The principal advantages are knowledge of the variety of problems that arose in the study that are not apparent in the final report, and the necessity of having given thought to various alternative designs that were not in fact used. The principal disadvantage lies in the necessity an author has to justify the work as it finally appeared. The disadvantage in this case may be reduced by the fact that I will use as a context for my examination a recent critical paper reviewing the report, "On the Value of Equality of Educational Opportunity as a Guide to Public Policy," by John F. Kain of Harvard University, and Eric A. Hanushek of the Air Force Academy. This is one of two papers written by economists critical of the Report, and includes most of the criticisms made by the other.* The paper by Kain and Hanushek, both economists, provides not only a check to the self-justification of an author but also the different perspective provided by a

---

*The other paper is Samuel Bowles and Henry Levin, "The Determinants of Scholastic Achievement--An Appraisal of Some Recent Evidence," Journal of Human Resources, Winter, 1968.

discipline that has been traditionally more closely linked to policy than has sociology, and one that has special perspectives of its own.

Defining the problem

The first question that arises in such a study as that requested by the Civil Rights Act in Section 402 is to determine precisely what the request means, and how it can be best fulfilled. In this case, the difficulty was especially great because the very concept of "equality of educational opportunity" is one that is presently undergoing change, and various members of government and of society have different conceptions of what such equality consists. There are many such conceptions and I will not go into them here, except to say that this was, as it should have been, regarded as the major problem in the design of the survey, and a great deal of attention was paid to it. A portion of an internal memorandum discussing the varieties of concepts of "equality of educational opportunity" has recently been published elsewhere, and I will not repeat them here.* It is sufficient to say that five were discussed: first, inequality defined by degree of racial segregation; second, inequality of resource inputs from the school system; third, inequality in "intangible" resources such as teacher morale; fourth, inequality of inputs as weighted

---

*James S. Coleman, "The Concept of Equality of Educational Opportunity," Harvard Educational Review, 38, Winter, 1968, pp. 7-22.

according to their effectiveness for achievement; and fifth, inequality of output as prima facie evidence of inequality of opportunity.

The study as designed and executed gave evidence relevant to all five of these definitions of educational opportunity. Kain and Hanushek argue that the most serious mistake of the study was here; that the study should have carried out a careful study of inputs, as the necessary minimum, before it could consider more esoteric questions, such as the effect of school inputs on achievement. Their charge is worth some discussion because if indeed a mistake was made at this point, it was the most serious of the study. But I believe that to have taken the approach proposed by Kain and Hanushek would have constituted exactly this magnitude of error.

As the survey was defined and carried out, it was intended to serve three purposes: to provide an accurate description of resource inputs for six different racial and ethnic groups at elementary and secondary school; to provide an accurate description of levels of achievement of each of these groups at three points in elementary school, grades 1, 3, and 6, and two in secondary school, grades 9 and 12; and to provide the basis for an analysis of the effects of various inputs on achievement. In terms of the five definitions of educational opportunity described above, such measures of effects were necessary for the fourth, to provide weights for various inputs, so that the "effective" inequality of opportunity could be assessed, and attention could be focused on those input resources that are effective in bringing about

educational opportunity, or by their unequal distribution, effective in maintaining inequality of opportunity.

As Kain and Hanushek point out, the sample design requirements and the kinds of measurement are different for each of these three purposes. In the first, the sampling variability is at the level of the school, even if reporting is ultimately to be done in terms of exposure of the average student to school resources, as the report did. The second and third aims, on the other hand, require measurements on students, in effect reducing the number of schools that can be included within the scope of such a study.* The third, analysis of the relation between input and output, imposes different design requirements than the second, in the way that analysis of relationships generally imposes a different sample design than does description of population characteristics, with less attention to sampling error, and more attention to the range of variability in the independent variables. Kain and Hanushek argue that the survey attempted too much; by attempting all three of these things, it failed to do well the first, minimum requirement.

This charge is a telling blow, for much of it is true. The final design is a compromise between three objectives, less good for any one of them than if the

---

*About 90% of the variance in student achievement lies within schools, so that the clustering effect that would be caused by sampling fewer schools and not sampling within schools (the latter of great administrative convenience) is not serious for measurement of achievement.

others had been absent. Its size is a compromise between measurement of school characteristics and of student characteristics, and its design a compromise between descriptive demands of the first two objectives and the analytical objectives of the third (for example, schools with intermediate proportions of Negroes and whites had especially high probabilities of being drawn). At one time in the survey design, in fact, a design involving two samples was seriously considered: a large sample of schools to measure school characteristics, and a smaller one for measuring student characteristics, including achievement, and for analyzing the relation of achievement to school characteristics. This design was rejected because the great effort necessary to secure cooperation of each school in releasing sensitive information would have made a much larger sample of schools difficult to achieve without sacrificing the other objectives.

The alternative, as proposed by Kain and Hanushek, was to do well the minimum necessary task: to measure carefully the input resources to schools attended by Negroes and those attended by whites, to show what is in fact the kind and degree of discrimination in schooling experienced by Negroes. They point to a number of specific weaknesses in such measurement, attributable to the more ambitious objectives.*

---

*I will not comment on these points in the text, but some comments on specifics are useful to correct misleading impressions some of the points may leave. First Kain and Hanushek point out the levels of nonresponse. They state that 61 percent of the high schools could not be included in the analysis, and of the high schools that were included, only 74 percent of their feeder elementary schools

The defect of the apparently simple and straightforward approach they suggest is the most serious possible: by selective attention to one of the definitions of equality of educational opportunity, that is, equality of inputs, it implicitly accepts and reinforces that definition. In effect, I suggest, it fails to see the

---

responded. Both these statements are in error. Seventy percent of high school principals completed a questionnaire, and seventy-four percent of the total number of feeder schools (based on the original sample of high schools) did so. Sixty-five percent of the schools returned student questionnaires (at 12th grade). However, in only 59 percent of the original high schools sample were matched student and principal questionnaires obtained, and the analysis was based on this sample.

This nonresponse of schools, together with item nonresponse on the questionnaires, is a problem that would have arisen in any of the research designs that might have been chosen, given the sensitivity of the problem. Serious biases may have been introduced, though a sample of nonresponding schools was drawn, and state records checked, showing little systematic differences on comparable items from the responding schools; but here as elsewhere, the question is whether to use data from a sample that may be biased, or to make policy decisions in the absence of data.

On the major points of nonresponse and miscoding mentioned by Kain and Hanushek (for example, nonresponse on principals' attitude questions, miscoding of school size by principal, poor coding of fathers' occupation) the problems were recognized by the staff, and the information not used in the analysis. School size, as used in the analysis, was obtained from the number of student questionnaires, because of the possible coding errors by principals, and the principals' attitude items were used only in the tabulations of school characteristics, where nonresponse was shown in the tabulation. Neither these attitude items nor father's occupation of child were used in the analysis of effects of school inputs on achievement. Thus the points made by Kain and Hanushek on these items are not relevant to the question of biases in the analysis.

forest because of too close attention to the trees. In contrast, the major virtue of the study as conceived and executed lay in the fact that it did not accept that definition, and by refusing to do so, has had its major impact in shifting policy attention from its traditional focus on comparison of inputs (the traditional measures of school quality used by school administrators: per pupil expenditure, class size, teacher salaries, age of building and equipment, and so on) to a focus on output, and the effectiveness of inputs for bringing about changes in output.

This effect of the study in shifting the focus of attention did not come about because the study gave selective emphasis to that definition of educational opportunity that entailed examination of effects; only one section of one chapter of the report was devoted to it. The study presented evidence relevant to all five of the definitions that had been initially laid out. It was the audience who, with evidence on all of these before it rather than only the comparisons of inputs that have traditionally served as the basis for comparisons of school "quality," focused its attention on the more relevant questions of output, and effect of inputs upon output. As I indicated above, I regard this shift of attention as the most important impact on policy of the study. It raised questions where none had been before: what is the value of the new large programs of federal aid to education? (The report results indicate very little, except through improving teacher quality, which the programs are not usually designed to do.) Do smaller class sizes bring increased achievement? (The report

results say no.) These and numerous other questions had been prematurely answered in the absence of facts, and if the study had taken the apparently straightforward careful approach that Kain and Hanushek propose, they would have continued to be answered prematurely, in the absence of facts. The study would have been celebrated for its careful accuracy, its measurement of inequality, and its irrelevance would have gone unnoticed, as policy-makers busily worked to eradicate those irrelevant inequalities.

I have spent so much attention upon this question of overall design because it is so important, and because one can be so easily misled. It appears most reasonable, from the standpoint of careful scientific inquiry, to limit policy-related research to that narrow definition of the problem that can give the most scientifically defensible results within the limits of time and resources available. But to do so may serve to define, and define incorrectly, the very policy questions that are addressed as a result of the research.

Inputs as disbursed and as received

Before turning to other questions concerning the validity of the survey's results in its description of inputs and achievement outputs, and in its analysis of the relation between them, it is necessary to discuss briefly a special problem that arises in the measurement of inputs of public resources to various groups in the population. This is a problem that will arise in other studies of such resource distribution, and it is well that it be discussed in some detail.

The problem arises from the fact that inputs can be viewed in two entirely different ways: inputs as disbursed by the school system, and inputs as received by the child. The difference can be shown by numerous examples: a school board can spend identical amounts on textbooks in two different schools (or two school boards can spend identical amounts in two different systems), so that the inputs as disbursed by school boards are identical. But if texts depreciate more rapidly, through loss and lack of care, in one school or one system than the other, then the text as received by a given child (say the second year after a new text is issued) constitutes a lesser input of educational resources to him than if he were in the other school or the other system. The examples could be multiplied endlessly: if teacher salary in a city and the surrounding suburban area are equal (and as the Report shows (Table 2.34.2) they are equal for schools attended by Negroes [largely in the central cities] and schools attended by whites in the same metropolitan areas [largely in the suburbs]), then the city is not competitive in salary, and loses the best teachers to the suburbs. Again, the inputs as disbursed by the school boards are equal, but the inputs as received by the children are not. As another example, if the expenditures on window glass in a city school in a lower class neighborhood and a suburban school were equal, the child in the city school would spend much of his time in classrooms with broken windows, while the child in the suburban school would not. Furthermore, nearly all the examples in which this "loss of input" occurs between disbursement and reception go in the same direction, that is, to reduce the resources received by the average Negro child.

The general principle can be described by an economic concept: the Negro child experiences external diseconomies through living in a lower class Negro neighborhood. (Sociologists often describe these as "contextual effects," but the fact that they represent real reduction of resources is better expressed through the term "diseconomies.") The fact that he himself may create external diseconomies for other Negro children is beside the point: those he experiences as a result of living where he does sharply reduce the resources he receives below those disbused by the school system.

Such a difference between inputs as disbursed and inputs as received creates enormous difficulties for any research designed to measure the "amount of resource input" from a governmental unit to any group in society. The fact that different external diseconomies are ordinarily highly correlated (e.g., the school that has frequently broken windows will be the same school that cannot hire the teachers it wants without special salary or other inducements) means that if inputs are measured as disbursed, this imparts a systematic bias to the measure if viewed as inputs received. Yet certainly from one point of view -- though not from all -- one is interested in input resources as received by the child.

It would be possible, of course, to make a virtue out of a fault -- to measure input resources as disbursed and received, so that one would obtain not only measures from both points of view but also by their difference, a measure of the amount of external diseconomy in each resource (for example, in teacher quality, in teaching materials, etc.) imposed on a child by living in a given

kind of neighborhood with a given group of schoolmates.

Yet to do this on a national basis would be an enormous undertaking, because of the difficulty of measuring resources as received, and would require a mixture of depth and comprehensiveness very difficult to achieve. For example, one resource never measured as an input resource is order and quiet in the classroom, presumably because it is a "free" resource. Yet one of the principal diseconomies some lower-class children impose on their classmates is the loss of this resource, the loss of order in the classroom. To measure the level of disorder carefully would be a difficult task. Another serious external diseconomy that lower class Negro children impose on others in their classrooms is to depress the level of teaching that a teacher can carry out in the classroom. Thus the teaching received by a child from a teacher in a lower class Negro classroom will be at a much lower level than that received from the same teacher in a classroom of middle class students performing at higher levels of achievement. Such a difference in inputs as received would be very difficult to measure. Or to measure the textbook resources as experienced by a child would require an intensive examination difficult to achieve on a national basis. Clearly it is important to measure the amount of diseconomy experienced by a child as a function of the kind of classmates and neighborhood, but it is an intensive analytical study that could hardly be carried out as part of a "simple and straightforward" study of equality of input resources on a national basis.

Thus even the apparently simple study of input resources becomes a rather complex one if it is viewed as it should be -- neither solely from the viewpoint of the administrator as distributor of resources, nor solely from the viewpoint of the child as recipient, but from the viewpoints of both.

It should be pointed out that this discrepancy between resources as disbursed and resources as received is and has been the cause of many disputes in the distribution of public resources generally. It can be obvious to a visitor to a ghetto school and a suburban school that the educational resources provided in the two are sharply different, ranging all the way from freshness of paint to the level of instruction in the classroom. But school administrators can then show that the same or greater resources are expended in the ghetto school than in the suburban one. The confused liberal (which many persons are on this question) often explains this as due to administrative juggling of figures to mask differences, and the administrator remains convinced he is right. He is right, but so is the observer who sees these sharp differences where the administrator says there are none.

This discrepancy between input resources as disbursed and as received is also very likely responsible for a large part of the confusion and disbelief attending the survey's finding of small differences or none between Negroes and whites for many input resources.* Many

---

*This unwillingness to accept the small degree of inequality of input resources is exhibited both in the paper by Kain and Hanushek and in the paper by Bowles and Levin. They cite several studies in particular cities;

observers "know" those inputs are different, but they know this by observations of the different schools, that is, inputs as received or experienced, not by examining the expenditures. The survey generally measured input resources as disbursed (that is, as reported by principal or superintendent) rather than as received, except in a few areas not ordinarily regarded as resources because they are not provided by the superintendent's office, such as the number of discipline problems reported by the teacher, the attitudes of teachers, and the educational backgrounds of a child's fellow students. These resources, incidentally, showed great differences between schools attended by the average Negro and those attended by the average white, suggesting the magnitude of the external diseconomy a Negro child experiences because of his neighborhood and classmates.

The analysis of school effects

Much of the paper by Kain and Hanushek is devoted to section 3.2 of the report, which carries out an analysis of effects of school resources upon verbal achievement. It is this section of the report that has occasioned much of the discussion surrounding it from persons concerned with school policy. This is as it should be, because as I have argued in the preceding pages, the question of effectiveness of school input resources

---

but many of these studies are marked by severe selective biases, since they aim to show how great the inequality of expenditures can be in selected cases, rather than how large it in fact is on the average.

is logically prior to the question of equality of particular inputs. To order things the other way around is reminiscent of the busy activity of southern school systems in constructing new buildings for Negro schools, increasing salaries of Negro schoolteachers, and buying new textbooks for Negro students in the period preceding the Supreme Court decision of 1954, to obtain an apparent equality of educational opportunity while leaving unexamined the question of whether these inputs were the important ones. As an aside, it seems to me likely that the ready acceptance by many whites of the policy of increasing the "quality" of all-black ghetto schools, whether advocated by white conservatives or by black militants, is similarly motivated: that this will solve the problem of Negro education without threatening the schools of the white suburbs. One might go so far as to say that the earliest cases of compensatory education for Negroes were the showcase Negro schools in the South of the early 1950's.

Thus the examination of effects of school factors was designed as a prior step to the description of "effective" inequalities of educational opportunity. I should go into the general design in a little detail, for it did not appear in the report as published, due to developments I will mention. The original intent was to carry out a regression analysis covering four general cluster of factors that might affect achievement: attributes of the child's own family background, characteristics of teachers, school resource inputs other than teachers, and social characteristics of the student body in his grade in school. The last of these

is described in the report as a cluster of student body factors, and they have been referred to elsewhere as peer factors. In effect, they are measures of some of the attributes of students in school that can exercise external economies or diseconomies upon the learning of a child in the school, through the addition or subtraction of "free" resources, and through the modification of input resources disbursed by the school administrators. The result of these regressions would then be two. The first result is the regression coefficients themselves, showing the relation of each of the teacher, school, and student body characteristics to achievement when all the other characteristics and family background were controlled. In unstandardized form, these regression coefficients would provide an estimate of the effect of one unit of the input resource on achievement, and in standardized form (e.g., as path coefficients), they would be measures of the relative importance of different factors in affecting achievement.*

---

*A note should be added here concerning what is meant by "achievement." Standardized tests, constructed by Educational Testing Service, were given in areas of verbal comprehension, non-verbal classification and analogy, reading comprehension, mathematics achievement, and at grades 9 and 12, five tests of specific subject areas. These test results correlated highly, and the one showing consistently higher correlations (both zero-order and partial) with school characteristics was the verbal comprehension test (taken from the SCAT series, principally a vocabulary test). This test was used throughout in the reported regression analyses. Regressions were carried out also on reading comprehension and mathematics achievement, and these showed similar results to the verbal achievement test, except that smaller proportions

The second result of the regression analysis was to be the principal one: these regression coefficients were to be used as weights for the various inputs, so that by replacing in the regression equation the levels of input resources for the average Negro in the region with those of the white in the same region, the predicted level of achievement would be changed. This would produce two results: First, a measure of effective inequality of opportunity would result as the increment in achievement that would be expected for the average Negro if all the input resources of schools and student bodies were at the level of those for the average white in the region. Second, by selectively changing in the equation some of the input resource levels to those held by whites, while keeping others at the levels held by Negroes, one could see which input inequalities were the effective ones, thus indicating which input resources would be expected to produce the largest effect if the input inequality for that resource were eliminated. The final form of this analysis was to be much like that of section 2 of the report, which showed differences in the levels of particular school input resources for the average Negro in a region, the average white living in the same county or metropolitan area, and the average white in the region and in the nation. The distinction of these tables would

---

of variance were explained. Thus achievement as I will use it in this discussion refers to vocabulary skills, but can stand also as a surrogate for achievement in the other areas mentioned.

be that the inequalities of resources, rather than being expressed by a difference in units of input resources (e.g., hundreds of dollars difference in teachers' salaries) would be expressed by the difference in existing average Negro achievement and predicted achievement if that input resource were at the level of whites in the same county or whites in that region or in the nation.

However, this plan was never carried through to final completion, but stopped short of the final step. The reason was collinearity among the various input factors, which I will discuss in more detail.

In doing this, I want to discuss several quite general problems that arose in this research and will arise in other research that attempts to assess effects of various input factors on some performance criterion. Some of these problems are directly related to the use that has been made of the report, and to criticisms that have been made of this use, or of the report itself, including that of Kain and Hanushek.

The first problem, well illustrated by this report and its interpretations, is the problem of determining exactly what is the policy question of interest, and then developing an appropriate statistical technique to give evidence regarding it. In this case, there has been widespread confusion about exactly this question, a confusion to which the report itself contributed. The analysis was designed to answer a single question: what is the relative importance for achievement of various resource inputs into schools, including the resources provided by other students (resources which I have earlier described as external economies or if negative,

diseconomies, imposed by other students). In the conceptual model we held of the student's performance, these school inputs, together with the child's own family background and his native ability which we regarded as unmeasurable, particularly in the absence of longitudinal data, constituted the principal determinants of motivation for and attitudes toward achievement, and this together with such motivation and attitudes, constituted the principal determinants of achievement itself. Now given this model, and given the policy interest in achievement, the overall relative effects of school factors on achievement (though not the mechanisms through which these effects occur) can be assessed by neglecting the intervening variable of motivation and attitudes. But given the differential degree of correlation of various school factors with the student's own background, and given the importance of these background factors for achievement, it is necessary to examine the relation of these school inputs to achievement when the student's own background is controlled. Otherwise, those school factors most highly correlated with the child's family background would show the strongest relation to achievement. The clearest case in which misleadingly high measures of effect would occur is for student body factors, because the backgrounds of other students are highly correlated with the student's own. Furthermore, these misleadingly high effects of school factors associated with the child's own background would <u>not</u> be sufficiently controlled by comparing sizes of standardized regression coefficients in a regression equation that includes family background

factors. The mathematics of regression analysis is such that when two variables are highly correlated and related to the dependent variable, then the multiple regression coefficients of both will include variance that is explainable by the other. However, another measure, $b^2(1-c^2)$, where b^2 is the standardized multiple regression coefficient of variable x_1 and c^2 is the correlation between the two independent variables, shows only the unique variance attributable to variable x_1.* Under such conditions, the following research procedure appeared most reasonable then, as it does now: to assess the relative importance of different school factors, given their differential correlation with the child's family background, the most accurate measure of relative importance is the additional variance in achievement that can be explained by the school factor, <u>after</u> family background factors have accounted for as much variance as they are able, that is, measures of the form of $b^2(1-c^2)$ rather than measures of the form of b. This gives an underestimate of the absolute effects of these school factors, insofar as they are responsible for

---

*Kain and Hanushek make an elementary error by describing this alternatively-explainable variance as due to the <u>interaction</u> of the two variables. That is not the case; in a linear regression equation, the variance uniquely explainable by the interaction of the two variables is not explained by either, while the variance under discussion here is that which, in a linear equation, is explainable by either, and ordinarily shows up in the regression coefficients of both.

some of the variance in achievement already explained by family background, but a better estimate of the relative effects than does the standardized regression coefficient.*

It is at this point that a confusion about the goals of the analysis arose. The way the results were reported contributed to this confusion, but was in no way misleading with regard to the policy conclusions that have been drawn from the report. Many persons, including Kain and Hanushek, have responded as if the goals of the analysis were to measure the relative effects of family background factors and school factors. But as I have indicated above, this was not the case at all, since policy alternatives concerned changes in various school factors (including student bodies, by distributing the external diseconomies imposed by lower class students among all students, through school integration), but not changes in the child's own family background.

As is evident in the discussion above of the technique used to assess relative effects, the technique would in fact have given misleading results if the goal had been different, for the relative effects of family background and school factors that it shows are biased in the direction of family background. Kain and Hanushek, among

---

*This statement must be qualified, because the latter half is true only under conditions that are not precisely known. I conjecture, however, that these conditions are that over half of the variance that could be alternatively accounted for by a given school factor or by family characteristics is in fact due to family. Results of numerous studies show that this condition is true for the case under consideration, that is school and family characteristics.

others, have mistaken the goals of the analysis, and have criticized the study for exactly this bias. But if the analysis had been carried out symmetrically, as had been initially intended, the goal of the analysis, and its relevance to the policy alternatives, would have been impaired. I will show shortly specific examples of how this might have occurred.

The results of the examination of relative importance of different school factors were that the most important cluster of factors was the social backgrounds of other students, the second most important was teachers' characteristics, and the lowest level of importance, explaining very little additional variance in most regions, was school facilities and curriculum characteristics. The policy questions, of course, are which school factors have more importance for achievement, and this ordering is the result of an analysis designed to answer that question. The most crucial policy issue is the issue of school and staff integration vs improvement of school facilities and curriculum while leaving unchanged the student bodies, and currently under black power pressure, leaving or bringing Negro teachers for Negro student bodies. It is worth remarking in this context that of all the teacher characteristics, those most highly and consistently associated with student achievement were two: the verbal skills of the teacher as measured by the score on a vocabulary test, and the racial composition of the teaching staff. These two were highly correlated, white teachers scoring consistently higher than Negro ones.

This relation of the racial composition of the staff is not apparent in the report, because it was entered in the regression after the verbal skills were entered, and under such conditions, explained little additional variance due to its correlation with them. Entering it in this order was based on the a priori assumption that if variance could be alternatively explained by teachers' verbal skills or teachers' race, the causal factor was more likely to be verbal skills than race. This was another example of the use of a priori assumptions rather than a wholly symmetric analysis. It was done, as in the case of family background, to prevent misleading inferences; but as in that case, interpretations should be made with knowledge of the asymmetry used in the analysis.

This high correlation between verbal skills and teachers' race, and their relation to student achievement, means that the policy alternative of improving facilities and programs of black schools and increasing the blackness of the teaching staff should, in terms of the report's results, reduce the achievement of Negro students, the one effect more than counterbalancing the other.

To return to the general point I wanted to make that has relevance for most policy-related research, this example of confusion about results shows clearly the importance of specifying the goal of the analysis in terms of the policy alternatives. The goal in this case affected even very technical points in the statistical analysis, and my conjecture is that it will do so in much policy-related research.

The confusion about the goals and results of the analysis has been increased in this case by several elements: first, the small amounts of additional variance accounted for by school facilities and curriculum led us as authors of the report to unduly focus attention upon the low absolute levels of additional variance explained, rather than solely upon the relative amounts explained by different schools -- although we did not make specific comparisons of family and school effects, because of the bias introduced by the asymmetric analysis; second, the interests of many persons in the report's audience other than those concerned with policy alternatives were in the question of the relative effects of family and school; and third, a very elementary confusion among some readers between effects of the child's own background (the size of which was not explicitly compared to effects of school factors), and effects of the social composition of the student body on a child's achievement, <u>apart</u> from his own family background, effects that were explicitly compared to school factors.

To show the misinterpretations that can arise due to technical errors when the policy questions are not kept clearly in mind requires examining in greater detail some of the results of the study. In doing this, I will present some further analysis carried out since the publication of the report.

In carrying out the regression analysis in the report, a technical reason in addition to the intellectual decisions discussed in an earlier section prevented the use of symmetric measures such as standardized regression coefficients for comparing the relative effects of different

school and student body factors on achievement. The result was not entirely satisfactory, because it entailed the comparison of added variances accounted for by school, teacher, and student body factors, entered in the regression equation in various orders. It is useful, then, to show symmetric measures of these factors and family background, all entered in the same multiple regression equation. The technical problem in doing so has now been solved, and the solution of the technical problems* allows illustration of the problems that arise by the direct use of the multiple regression coefficients.

Two tables are presented below, both containing symmetric measures taken from multiple regression equations in which all four clusters of variables are entered.

---

*The technical problem in short was this: it was desired to get a measure of the overall relation of the cluster of teacher characteristics, the cluster of school facilities and curriculum characteristics, the cluster of student body characteristics, and the cluster of characteristics of the child's own family background. But it is not possible to do this by adding the multiple regression coefficients for all variables within a cluster, because the sum would be too large, containing the same variance included in the regression coefficients for several different correlated variables. The solution is to use the regression coefficients from the total equation including all variables, as weights in forming four new composite variables representing the four clusters mentioned above. Then a regression of achievement on these four composite variables gives a standardized regression coefficient for each of the four clusters. In the new regression equation with four independent variables, rather than 29, exactly the same amount of variance in achievement is explained; it merely allows one to obtain a standardized regression coefficient for each of the clusters.

These clusters were each entered as a single variable, an index which used as weights the multiple regression coefficients on the individual variables within the cluster, so that the total variance accounted for is the same as in the original equation, but unlike that equation, a single standardized regression coefficient for the cluster is obtained. Table 1 contains these standardized regression coefficients (which can also be regarded as path coefficients). Table 2, however, contains measures of the sort that were used in the report, except that these measures are presented for all four of the clusters at issue: the student's own background, school curriculum and facilities, teacher characteristics and characteristics of the student body. The measures are measures of the unique contribution to variance in verbal achievement, after all three of the other clusters of variables are entered in the equation. These unique contributions have been scaled up so that their sum equals the total variance explained (the square of the multiple correlation coefficient).

Table 3 summarizes these two tables over grades and regions, for Negroes and whites separately, to facilitate comparison.

The major difference between these two measures is that they "control" on the other variables in different senses. When two independent variables are correlated, then the variance that may be explained by either contributes to the regression coefficients of both. In using the variance uniquely explainable by a variable, however, the variance explainable by either is not allocated to

either variable. Thus the regression coefficients give a liberal estimate of the effect of each, and the unique contributions to the variance give a conservative measure. The question in using one or the other for purposes such as this, however, is to get a good estimate of the relative effects of each cluster of variables (in this study, an estimate of the relative effects of the three school-related clusters).

The problem that can arise by using regression coefficients is well illustrated by the coefficients for family characteristics and student body characteristics. Among Negroes in the South, at every grade level, the regression coefficient for study body characteristics (which is correlated with the child's own family background) is comparable in size to the coefficient for family background, and in grade 12, it is even higher than that of family background. The unique contributions to variance in table 2, however, show that in no cases is the unique contribution of student body characteristics near that of family background. In grade 12, where the multiple regression coefficients are .23 and .22 for student body and the child's own family, the unique contributions are .078 and .119.

It is in cases like this where the use of multiple regression coefficients can be misleading. If such coefficients had been presented for the four sets of variables, then it would have led to the conclusion that in the South, the characteristics of the student body in the school are as important for a child's achievement as is his own background, a conclusion that appears false

on its face and a conclusion that is not drawn from the relative sizes of the unique variance contributions.

It is paradoxical that the objections to the report's use of unique variance contributions rather than regression coefficients have been by the two pairs of economists, Bowles and Levin, and Kain and Hanushek, both of whom objected also that the use of regression coefficients would have shown greater school and teacher effects relative to the student body effects. But as comparison of these two tables shows, the variable whose apparent effect is most reduced by examining unique contributions rather than multiple regression coefficients is the student body characteristics.*

---

*Kain and Hanushek carry out a regression analysis which purports to cast doubt on our conclusions, but does so by leaving out two of the three most important student body variables (school attendance and proportion planning to attend college) and substituting others in their place. Their example is presented as an illustration, a "representative case" among "several different models" they analyzed. But it could hardly be more misleading if it were intended to mislead. The table is presented here with their results and with the results obtained from using those variables that were used in the report. I have not run regressions in every order, but those that have been run are sufficient to show the misleading nature of their "representative case." They appear also to have made an error which makes their school contribution 1.5 times to 1.7 times its actual value. The case, incidentally, is hardly representative, since school characteristics are negatively correlated with other variables, producing the increase in unique contribution shown in comparing the figures in parentheses for the third and fifth columns for family characteristics and the figures outside parentheses for the third and fifth columns on school characteristics.

PROCESSING CONTROL RECORD

for

MUST and MUST - EXPEDITE DOCUMENTS

ERIC FACILITY

Document Identification Number: BR-6-1610-25

Date Received from Central ERIC: JAN 1 1969

Date Shipped to Clearinghouse: JAN 1[illegible] 1969

ERIC CLEARINGHOUSE

Date Received: JAN 30 1969

Clearinghouse Accession Number: EA 001 956

Date Shipped: 4-18-69

ERIC FACILITY

Date Received: April 24, 1969

ED Accession Number: _____

Date Shipped to EDRS: _____

However, apart from the extreme cases examplified by the Southern Negroes, the regression coefficient and unique contributions do not give radically different results. Table 3 shows that by both measures, family background is clearly the strongest cluster, with student body characteristics following for Negroes, and both school and teacher characteristics following that, while for whites all three of the latter are smaller, and all three are nearly alike. The similarity of these two measures is more apparent if it is recalled that the unique variance is a measure that should be compared to the square of the regression coefficients. If the regression coefficients in table 3 are squared, they are much more comparable in relative magnitudes to the unique contributions.

The reason for the rather good comparability between the square of the regression coefficients and the unique contributions for most of the grade-region-race groups in tables 1 and 2 is shown in table 4, which gives an indication of just how highly correlated these four

---

Table 1 from Kain and Hanushek, together with comparable figures in parentheses, using variables from the report.
Added proportion of variance explained

Variable Cluster	Order of entry: 1	2 after S	2 after P	2 after F	3
S (Teacher, facilities curriculum)	(.0466) .0808		.0222	(.0363) .0560	(.0185) .0312
P (student body)	(.0788) .0703	(.0487) .0117		.0420	(.0349) .0072
F (family)	(.0748) .0777	.0529	(.0487) .0494		(.0505) .0484

clusters of variables are for each grade-region-race group. Table 4 gives the total variance explained and the sum of the unique contributions to this variance. If the four clusters of variables were uncorrelated, the two numbers would be the same, and the squared multiple regression coefficients would equal the unique contributions. If they were all perfectly correlated, the unique contributions would be zero.

As Table 4 shows, it is for Negroes in the South where the clusters of variables are most highly correlated. Among Negroes in the North, there is very little correlation among them.*

Multiple modes of analysis

Another quite general point I want to make about policy-related research is the importance of using different modes of analysis to examine the same question. If these modes are mutually confirming, the results are considerably strengthened; if not, they are weakened. In this case, the example I want to use is the question of the absolute magnitude of school effects, for in articles

---

*Both Kain and Hanushek and Bowles and Levin make the elementary mistake of discussing collinearity in the original matrix of 29 independent variables as if it were the collinearity among the four clusters of variables. Both pairs miss the essential point that it is only collinearity between the clusters of variables that matters in these interpretations. Collinearity between two teacher variables, for example, has no effect upon the collinearity between clusters, and thus none on the interpretations made here, as it should not, while it has an effect on the collinearity found in the matrix of the original 29 variables.

subsequent to the report, I have used evidence from the study to draw strong policy conclusions from the absolute level, arguing that the low absolute level of effect means that a more radical modification of a child's environment than that provided by schools is necessary to induce achievement in children whose family environments do not insure learning.

In the present research, it would have been incorrect to base such arguments principally on the results of the regression analysis, because the techniques used there were designed to more accurately assess relative effects of school factors at the risk of underestimating absolute size of effects. For this reason, it is especially important to have several alternative grounds for such inference. And it is useful to have these other sources of evidence based on analyses as different as possible in form, so that the errors of one will not appear in the other. There were three sources of such evidence, supported by the results of the regression analysis:

1. In an analysis of variance, the generally low proportion of variance in achievement that lay between schools, for each racial group, between 15 and 20% for Negroes, and less than 15% for whites. This means that the major portion of the variance in achievement could never be accounted for by differences between schools, for it resides within the school itself. If schools had strong and differential impacts on achievement (and the size of the differences in impact can be expected to be proportional to the strength of impact), then children within a given school should be achieving more nearly at the same level than the study showed to be the case.

2. This analysis of variance was carried out for each of the grade levels, grades 1, 3, 6, 9, and 12. If schools have strong and differential impacts, then the proportion of variance between schools should change over the school years. The between-school variance at the beginning of grade 1 is merely due to the differences between differing entering student bodies, due solely to family backgrounds, except for variance due to the test-taking situation in the school. If school effects are strong and positively correlated with family background, as all evidence would suggest, the between-school component of variance should <u>increase</u> over the years of school. If they are strong and uncorrelated with family background, the between-school component of variance should <u>decrease</u> over the years of school, or perhaps first decrease as student input differences are washed out, and then increase as school differences make student bodies diverge in achievement.

But as it turns out (Table 3.22.1), very little happens to the between-school component of variance. It remains about the same over the years of school. The simplest explanation of this is that the initial differences with which children enter school simply continue over the years of school, unaffected by the impacts of good or poor schools. The kind of school influences that would produce such a result are those that merely carry children along at the same relative levels

of performance with which they begin school.*

3. The correlation between family background and achievement is approximately constant over grades 6 to 12 for both Negroes and whites (Table 3.221.3). (For grades 1 and 3 it is lower than for the later years, but this may be due to poorer measurement of background at these levels.) The absence of a steady decline in this relation over time indicates that schools do not constitute an important enough modification of the child's environment to interrupt the family processes that in the absence of school would be expected to show the same constant correlation with achievement that they now show.**

These three modes of analysis, reinforced by the regression analyses that show low unique variance explained by school factors (with the exception of some teachers' characteristics), provide a rather strong base of evidence for the inference that school factors constitute a

---

*Various persons have argued that this and other results of the study simply show the importance of fixed genetic differences among children. Such genetic differences, within and between schools, could produce the observed constancy of between-school component of variance, but only if school effects were relatively weak or uniform. If school effects were strong, and different for different schools, they would magnify the between-school component of variance through interaction with the genetic factors.

**For some of the ethnic groups other than Negroes and whites, a decline does occur between grades 6 and 12. For technical reasons, however, less confidence should be placed in those regressions than in those for Negroes and whites.

relatively minor modification of the child's learning environment, a strong basis for the argument that much more radical modifications of the environment are necessary in order to greatly increase achievement of presently low-achieving groups.

This result illustrates my more general point about policy-related research -- that it should obtain evidence from analyses as technically different as possible, to strengthen the grounds for inference. It is particularly important in this case that two of the results (#2 and #3) were based on comparison over grade levels from grade 1 to grade 12, since the regression analyses were necessarily carried out within the same grade level, and the inferences about small school effects were not based on trends over different grades.

The inadequacy of general field surveys for answering specific policy questions

The problems of interpretation of results in a massive study like this one illustrate another general point in policy research. This is the inadequacy of analysis of a general sample of institutions or students for answering very specific questions relevant to policy, when the policy-related variables of policy interest are highly correlated, and have relatively small effect on the dependent variable under study. The results of this survey show only the most general outlines of the factors affecting school achievement. For answering specific questions, it is clear that methods are necessary which empirically rather than analytically separate out the

variables of policy interest. One way in which this can be done is, of course, through experimental research. This, however, has the defect that the effects occur over time, and experimental research must involve time in which the experimental variables can have their effect.

In the absence of the necessary time, it appears likely that other methods are possible, if one recognizes certain dangers inherent in them. One of the most appealing for a study like this, in which much data beyond that necessary for analysis is obtained, is the use of computer procedures for selecting students within schools for which certain input variables are orthogonal and others are perfectly correlated. Regression analysis on those orthogonal sets of variables will provide estimates of the effects of the sets of perfectly correlated variables. Then further computer selection can be used to identify students for which the variables perfectly correlated in the first analysis are orthogonal, and an analysis of these students used to examine the relative effects of the newly orthogonal variables.

The details of such a procedure, the statistical problems it might introduce, and the methods for reducing these problems, are not clear. It is evident, however, that since the samples are not representative of the population of students of interest, the parameters estimated are specific to the students thus sampled, and may be in considerable error for the population of interest. Even so, some procedures in which estimates made on the samples resulting from search were used for prediction in the unbiased sample and inferences about

biases made from the errors of prediction, seems possible.

In short, it is quite evident that much work remains to be done in devising techniques that can, within reasonable time constraints imposed by policy problems, give better estimates of the expected effects of policy changes.

One aid to this, suggested by Kain and Hanushek in their critique, is the development of more fully-elaborated conceptual models for use in the statistical analysis. For example, as they point out, school effects are, or should be, cumulative over time. As a consequence, the appropriate model and measurement should involve the product of the school resource times the length of time to which the child has been exposed to it. If all children remained in the same schools, were subject to the same home environmental conditions, and were in schools with the same kinds of other students, throughout their school lives, then time is unnecessary in the model. However, if he has moved and if his peers in school have changed, both of which are true for some students, then resources have been available for different amounts of time, and time should be explicitly incorporated into the analysis.

Still another approach to these policy questions might be to examine students at different grade levels in the same schools, controlling on family backgrounds to "standardize" the student body at one level against that at another. The dependent variable in this case would be the difference in achievement levels of standardized student bodies at two grades, or the inputed "growth rates." These "growth rates" can then be related to the characteristics of the school.

For these data, at least two such analyses have been carried out, one by David Armor at Harvard, and another by me. Armor's analysis generally confirms that of the report, but as the report, is unable to show effects of more specific school variables of policy interest. My analysis is not complete, but at present appears not to show any new results.

More generally, it appears that the most promising possibility for policy research lies in much more systematic and careful administrative records of social institutions. These records, if they were well-maintained and comparable among schools (or for other policy questions, among other institutions), would allow analyses for policy questions to be carried out regularly and at minimal cost, by local school systems, by state systems, or nationally.

* * *

Altogether, it is clear that research to examine questions of policy can be done to provide a better base for general directions of policy. I believe Equality of Educational Opportunity has done so principally through the way in which the problem was defined, resulting in a redirection of attention from school inputs as prima facie measures of quality to school outputs, and resulting as well in an expansion of the conception of school inputs beyond those intentionally supplied by the school board. It is equally clear, however, that policy research in social areas is only beginning, and that social scientists have much to learn about how to answer policy-related research questions.

Table 1

Standardized multiple regression coefficients (or path coefficients) as measures of the importance of each of four clusters of variables on verbal achievement, at grades 12, 9, and 6. Family background (six variables), school facilities and curriculum (11 variables), teacher characteristics (7 variables), and student body characteristics (5 variables). Negroes and Whites in North and South, grades 12, 9, 6.

	Negro North	Negro South	White North	White South
Grade 12				
Family	.23	.22	.34	.34
Facilities & curriculum	.13	.07	.10	.07
Teacher	.13	.12	.09	.04
Student Body	.23	.23	.09	.11
R^2	.15	.23	.16	.17
Grade 9				
Family	.26	.22	.40	.38
Facilities & curriculum	.14	.16	.10	.05
Teacher	.12	.09	.11	.09
Student Body	.16	.19	.08	.07
R^2	.12	.21	.19	.20
Grade 6				
Family	.27	.29	.34	.41
Facilities & curriculum	.04	.14	.05	.06
Teacher	.14	.12	.07	.10
Student Body	.14	.12	.12	.07
R^2	.13	.21	.16	.20

Table 2

Unique contributions to variance in verbal achievement (scaled up to sum to R^2 in each regression) as measures of the importance of each of four clusters of variables: family background (six variables), school facilities and curriculum (11 variables), teacher characteristics (7 variables), and student body characteristics (5 variables). Negroes and Whites in North and South, grades 12, 9, 6.

	Negro North	Negro South	White North	White South
Grade 12				
Family	.067	.119	.133	.144
Facilities & curriculum	.018	.009	.014	.007
Teacher	.016	.026	.009	.002
Student Body	.046	.078	.008	.013
R^2	.146	.232	.165	.166
Grade 9				
Family	.065	.098	.160	.183
Facilities & curriculum	.017	.046	.010	.004
Teacher	.012	.014	.010	.010
Student Body	.027	.053	.006	.005
R^2	.121	.211	.186	.203
Grade 6				
Family	.086	.137	.135	.179
Facilities & curriculum	.002	.033	.004	.003
Teacher	.024	.023	.007	.011
Student Body	.021	.021	.017	.006
R^2	.132	.213	.163	.199

Table 3

Averages of standardized multiple regression coefficients (from Table 1) and unique contributions to variance (from Table 2) for Negroes and Whites separately, averaged over grades 12, 9, and 6.

	average of standardized regression coefficients		average of re-scaled, unique variance contributions	
	Negro	White	Negro	White
Family	.25	.37	.095	.156
Facilities & curriculum	.11	.07	.019	.007
Teacher	.12	.08	.021	.008
Student Body	.18	.09	.041	.009

Table 4

Total variance explained (r^2) and the sum of unique contributions to the variance, for each grade, race, and regional group.

	Negro North	Negro South	White North	White South
Grade 12				
r^2	.146	.232	.165	.166
sum of unique contributions	.111	.080	.127	.121
Grade 9				
r^2	.121	.211	.186	.203
sum of unique contributions	.111	.097	.179	.146
Grade 6				
r^2	.132	.213	.163	.199
sum of unique contributions	.108	.115	.119	.180