DOCUMENT RESUME

ED 059 174 SP 005 541

AUTHOR	Kiesling, Herbert J.
TITLE	Input and Output in California Compensatory Educational Projects.
SPONS AGENCY	Carnegie Corp. of New York, N.Y.; Rand Corp., Santa Monica, Calif.
REPORT NO	R-781-CC-RC
PUB DATE	Oct 71
NOTE	61p.
EDRS PRICE	MF-$0.65 HC-$3.29
DESCRIPTORS	*Compensatory Education; *Learning Processes; *Program Evaluation; *Reading Improvement; *Reading Programs
IDENTIFIERS	*Stanford Reading Test

ABSTRACT

This report describes a study of the relationship of instructional process and program organization to pupils' learning in Title I compensatory education projects, as measured by the Stanford Reading Test. This is the first attempt to apply economic input/output methodology to compensatory education. Personnel in 42 projects in 37 California school districts were interviewed to obtain detailed data on teaching strategies, individual instruction time per pupil, intensity of instruction, patterns of coordination of project personnel, and other variables. Variables were related to pupils' monthly gain in grade equivalents via multiregression techniques, holding program length and beginning score constant. Results contradict reports that compensatory education is ineffective. Individual instruction by trained reading specialists was consistently related to gains. Less strongly related were staff planning time and individual instruction by classroom aides. The six best projects averaged at least 1.25 months' learning per month of instruction. None were large or urban, all had small group instruction by specialists, high ratio of managers to pupils, and planning coordination. (MBM)

SP
N-CO

R-781-CC/RC

October 1971

Input and Output in California Compensatory Education Projects

Herbert J. Kiesling

Rand
SANTA MONICA, CA. 90406

R-781-CC/RC

October 1971

Input and Output in California Compensatory Education Projects

Herbert J. Kiesling

This research was jointly sponsored by a Grant from the Carnegie Corporation of New York and by The Rand Corporation. Views or conclusions contained in this study should not be interpreted as the official opinion or policy of the Carnegie Corporation or of The Rand Corporation.

Rand
SANTA MONICA, CA. 90406

Bibliographies of Selected Rand Publications

Rand maintains a number of special subject bibliographies containing abstracts of Rand publications in fields of wide current interest. The following bibliographies are available upon request:

Aerodynamics • Africa • Arms Control • Civil Defense
Combinatorics • Communication Satellites • Communication Systems
Communist China • Computer Simulation • Computing Technology
Decisionmaking • Delphi • East-West Trade • Education
Game Theory • Health-related Research • Human Resources
Latin America • Linguistics • Maintenance • Middle East
Policy Sciences • Pollution • Population
Privacy in the Computer Age • Program Budgeting • Public Safety
SIMSCRIPT and Its Applications • Southeast Asia
Space Technology and Planning • Statistics • Systems Analysis
Television • Transportation • Urban Problems • USSR/East Europe
Water Resources • Weapon Systems Acquisition
Weather Forecasting and Control

To obtain copies of these bibliographies, and to receive information on how to obtain copies of individual publications, write to: Communications Department, Rand, 1700 Main Street, Santa Monica, California 90406.

Published by The Rand Corporation

PREFACE AND ACKNOWLEDGMENTS

The research reported in this study was jointly sponsored by the Carnegie Corporation of New York and The Rand Corporation. The manuscript as it now stands benefited from the careful readings and comments of Rand colleagues Kent Anderson, Stephen Barro, Margaret Carpenter, Stephen Carroll; Thomas Ribich of the University of North Carolina; and Burton Weisbrod of the University of Wisconsin. The author also wishes to express appreciation to Harvey Averch, Ted Donaldson, and Marjorie Rapp of Rand for helpful advice, and profound gratitude to the Title I managers and other California school personnel who virtually without exception granted him their full cooperation. Although the school district personnel must remain anonymous, persons who assisted from the California State Division of Compensatory Education are Thomas Shellhammer, Alex Law, William McCormack, and Gerald Ryder, and from the Los Angeles County Department of Compensatory Education, William Turner and Norma Wilbur. The author is a member of the Economics Department of Indiana University and a consultant to The Rand Corporation.

SUMMARY

This Report describes a study of the relationship of instruction process and program organization to the gain of pupils in California compensatory education projects in the Stanford Reading Test. The methodology follows the "input-output" or "production function" approach of the economist.

Personnel in 42 projects in 37 school districts were interviewed to obtain detailed data on teaching strategies, intensity of instruction, patterns of coordination of project personnel, and other variables. Variables were constructed from these data and related to gain per month in grade equivalents holding the effects of program length and beginning score constant.

The findings were that the amount of instruction given by trained reading specialists is consistently related to pupil gains. There was some evidence to show that planning time and instruction by paraprofessional teaching personnel aiding the regular classroom teacher were also related to gains.

CONTENTS

I. INTRODUCTION

PURPOSE OF THE STUDY

The study described in this Report is intended to advance our knowledge of compensatory education, especially with respect to issues of program design relevant to the allocation of educational resources. A second, equally important objective is that the study will also prove valuable as a methodological experiment. The massive effort to overcome educational handicaps due to cultural deprivations authorized by Title I of the Elementary and Secondary Education Act of 1965 is one of the more important national social innovations of recent years. The program is costly, financed at an average of more than $1 billion annually; and it is broad, aimed at all children coming from families officially classified as being "poor."[1] Sponsors and proponents of the legislation have placed high hopes upon the measure as being one way of pulling alienated poor and minority children into the mainstream of American life.

Despite its obvious importance, the program has been extremely difficult to evaluate, largely because no research methodology has been developed whose results were useful to the policymaker. Any study that makes a contribution to our knowledge of the substantive questions must break new ground in the methodology of educational research as well. Such an effort is made in this study.

ORGANIZATION OF THE REPORT

Because of the methodological interests just discussed, Section II includes a discussion of the place of this study in policy relevant education research. The following section deals with the steps taken to derive a model of compensatory education. It includes a description of past findings that suggest hypotheses to test, a description of the

[1] The Report on Title I for the 1968 fiscal year gives the number of children in poor families as 7,700,000 [31, p. 66]. Of these, 89 percent are in schools that receive Title I aid and about 52 percent are participating in some form of Title I program [31, pp. 14, 87].

compensatory education process (which is used to generate testable hypotheses), and a discussion of the variables collected by questionnaire. Section IV contains the findings, and Section V is devoted to a discussion of implications of the study for further research.

II. METHODOLOGICAL CONSIDERATIONS

GENERAL BACKGROUND

In the past there have been two fundamental approaches to policy relevant evaluative research in education. To use descriptive terms developed by Averch et al. [2], they are the "process" and "input-output" approaches. The process approach, which characterizes most past educational research, is usually done in carefully designed experiments, often using experimental- versus control-group methodology. These studies tend to have no standard method for reporting such student characteristics as socioeconomic background, attitudinal variables, and the like (beyond merely ascertaining that such characteristics are the same for both experimental and control groups). The criterion measure, or measure of performance, is whatever the researcher chooses, and there is very little consistency from study to study in terms of criterion measures, or if there is, the measures usually are of little direct interest to policymakers.[1]

In the input-output approach quantifiable output measures, such as standardized objective test scores, are related to quantities of resource inputs, with some care being taken to account, at least roughly, for student differences in learning rate due to socioeconomic characteristics. This methodology overcomes the basic weaknesses of the process approach by using large samples with the same measure of output, but at the same time it lacks the basic strength of process studies, which is the student-specific (or at least classroom-specific) nature of the analysis. The variables used have been aggregated by school buildings or school district (often for just one grade) and, further, they have not measured the personal traits of teachers or other school personnel but what Stephen Michelson has aptly termed their "objectified

---

[1] For example, many of the criterion measures of teacher performance are ratings by their superiors as to the quality of their performance. There is seldom any effort to obtain correlations of ratings by superiors and actual classroom performance.

characteristics," years of experience, number of degrees, and the like.[1]

An important difference between the process and input-output methodological approaches is the statistical techniques they normally employ. Well conducted process studies have traditionally compared the means of treatment and nontreatment groups for statistically significant differences. The emphasis has been upon finding that one treatment yields results that are "better than" another, without focusing greatly upon how much better the treatment group performed. Input-output studies have, on the other hand, used multiple regression techniques which, if assumptions underlying the statistical analysis are reasonably satisfied, have the important advantages of (1) being able to trace functional relationships between variables, and (2) to do so net of the effects of other variables. These advantages make the approach potentially a more powerful statistical tool than the analysis of variance designs used in process research, although the latter are somewhat superior perhaps for studying interaction effects.

---

[1]Two exceptions to these remarks must be noted. One Rand-sponsored study by Hanushek [15] has matched pupils in grades 2 and 3 with their teachers. Also a number of studies, including those based on the Coleman report and the Hanushek study just mentioned, have had variables for teacher performance on a simple verbal abilities test.

III. BUILDING A MODEL OF THE COMPENSATORY EDUCATION PROCESS

DESCRIPTION OF THE COMPENSATORY EDUCATION PROCESS

The model used in this report is based upon a descriptive analysis of the compensatory education process and upon the findings of earlier studies of compensatory education programs. To begin constructing a model of compensatory education it is useful to identify meaningful input variables through detailed analysis of the process sequence. In constructing the model for the empirical analysis, therefore, the starting point was a careful consideration of the problem of educating each child, including the organization, preparation, and actions that must be undertaken by the school in dealing with this problem from beginning to end.

In general, the "problem" of education usually begins with the realization that the pupil does not possess skills and attitudes society wishes him to have. The education process deals with the "problem" of lack of knowledge. A strategy for doing this includes the training of instructional personnel, the planning of instruction, and the testing of results. In most traditional American education, preparation of instructional personnel occurs at the university, while planning and testing is the function of the individual teacher who is not supervised to any great extent.

The education problem for children who are seriously underachieving should be viewed somewhat differently from that for normal children. Instead of "normal" lack of knowledge there is an "abnormal" lack of knowledge, implying some special reason for it; and the discovery of such reasons (diagnosis) becomes the important first step of compensatory education. Whether done explicitly or tacitly, formally or informally, the education of underachievers must begin with successful program diagnosis as a part of Title I programs.

Successful diagnosis immediately implies the need for proper prescription of instructional techniques to deal effectively with the problems found in the diagnosis, and the second step in the process is, therefore, prescription.

The third step is to communicate the prescription for successfully overcoming the problem to instructional personnel, who, along with program managers and other decisionmakers, must design and implement instructional techniques, the fourth step. The fifth and final step is to evaluate the success of the program. The evaluation step, especially if there is experimentation with different techniques, provides important feedback to all the other steps in the process.[1]

Although it is conceivable that a compensatory education program could get by without coordination of project members and effective leadership by the project director (for example, in a project completely run by a reading specialist), in almost all instances I observed, teamwork of project personnel has been important. Thus, even when a specialist is in complete control of a program, it appears desirable that she communicate regularly with the classroom teachers of the children.

PRIOR FINDINGS

The research findings of two prior studies provide useful information about which aspects of the process just described should be contained in an input-output model. One is an earlier telephone interview study of projects that were described by California State Compensatory Education personnel as highly successful.[2] Project directors were asked to describe their projects and to point out features they considered central to program success. The second set of studies was the painstaking review of project evaluations done by Hawkridge and a number of associates at the American Institutes for Research.[3] The authors first described the characteristics of studies they could pinpoint as being successful. Then they found a number of projects that were quite similar to the successful ones in terms of

---

[1] See Rapp [28].

[2] The success criterion used was gains in cognitive reading tests that approached two times what was considered "average" for low SES children. See Kiesling [23].

[3] [16, 17, 18].

objectives, basic program type, and pupil age, and they attempted to ascertain which program components were associated with success and which with failure.

The findings for the studies just mentioned are briefly summarized in Table 1. Well planned, individualized instruction appears to be the key attribute of successful programs. Good in-service training is given prominent mention as well. Hawkridge and his associates also concluded that motivation by pupils' parents was also important, at least at the elementary school level. These become, then, the program aspects that should be traced with special care in the analysis. In the next few pages the operation of compensatory education programs is considered in somewhat more detail in order to help in deriving workable variables.

INDIVIDUALIZED INSTRUCTION[1]

For purposes of this study, instruction can be divided into group and individualized techniques. In group instruction all members of the class encounter the same set of experiences: they hear the same teacher lectures and comments by their peers, participate in the same exercises, and so forth. Students are required to learn at some minimum rate which is the same for everyone, although upward departures from the minimum are encouraged and rewarded.

When instruction is individualized, there is a relationship or interaction of the instructor directly with the individual pupils. Assignments are based on the individual needs of the student according to his ability, motivation, learning habits, previous attainments, and so forth. Sometimes pupils are given a degree of choice concerning curriculum in light of their own goals. Individualized instruction always involves individual diagnosis and testing to ascertain the pupil's problems and strengths. Sophisticated diagnosis may suggest the kind of instructional techniques that might best be used for

[1]The following discussion has benefitted greatly from the series of monographs on the subject of individualized instruction written at the Far Western Regional Laboratory [9].

Table 1

FACTORS ASSOCIATED WITH SUCCESSFUL COMPENSATORY EDUCATION PROJECTS ACCORDING TO STUDIES BY HAWKRIDGE AND KIESLING

Hawkridge

Pre-school Programs

1. Careful planning, including statement of objectives
2. Teacher training in the methods of the program
3. Instruction and material closely relevant to the objectives

Elementary Programs

1. Academic objectives clearly stated
2. Active parental involvement, particularly as motivators
3. Individual attention for pupils' learning problems
4. High intensity of treatment

Secondary Programs

1. Academic objectives clearly stated
2. Individualization of instruction

Kiesling

1. Individualization of instruction
2. Thorough planning and program coordination
3. Thorough in-service training of teaching personnel

Sources: [18], pp. 19-20; [23], p. 8.

each child, or this may be ascertained in the course of instruction with experimentation. Pupil progress is evaluated continually.[1]

Although individualized instruction is a complex process, this report will focus upon three key features that are central to its working: the intensity of instruction, or the amount of instruction given to the pupil; the types of personnel and methods used to deliver the instruction to the pupil; and the type of instructional materials used.

Instructional Intensity

It is reasonable to expect that the amount of instruction given to pupils, other things being equal, would make a difference to program success. It is necessary to account for four sources of variations in treatment in measuring intensity: (1) the number of minutes per day each child is seen, (2) the number of instructional sessions per week the child has, (3) the number of teaching personnel working with him, and (4) the number of pupils receiving instruction.

Instructional Design

American public schools have considerably more variation in the design of instruction for compensatory education than for normal education. Three kinds of personnel may be used in compensatory education: the regular classroom teacher who is released from part of her duties so she can give additional instruction to the compensatory education child; the trained specialist; and the paraprofessional, who is enlisted in support of either classroom teachers or specialists.

---

[1] Despite what may seem logical in the matter, class size for individualized instruction is not necessarily smaller than that for group instruction. It is the teaching technique, not the class size, that is important. Group instruction, with virtually no individualized instruction at all, could be carried on (and often is, for example, in graduate courses) with classes of four or five. Individual instruction techniques often include giving the child a short assignment and sending him off to do it. A good specialist instructor can probably give individualized instruction to 20 children at once.

(Paraprofessionals are instructional personnel who are given on-the-job training and who do not have the required levels of formal education normally required for certification as a classroom teacher or as a specialist.) Also, the instruction itself is given either in the regular classroom or in some separate facility, usually a resource facility equipped with special materials and supplies.

Since specialists receive training in individualized instruction techniques, use of such personnel should yield better results. Guszak [12] concludes that the disadvantaged child is best taught language skills by a diagnostic reading teacher who understands the variety of reading skills that exist and who can tailor instruction in skills to the individual while providing him with the emotional support that makes him wish to work and achieve. Guszak also suggests that "the rank and file of teachers do not possess systematic knowledge of their reading skills program." [12, p. 363].

In light of the many criticisms of the role of certification in teaching effectiveness that have appeared in recent years,[1] it is also of great interest to analyze the role of the paraprofessional in the instructional process.

Instructional Materials

The type of instructional materials used will very likely make a difference in the effectiveness of individualized instruction. Materials and equipment that are commonly used in much greater depth for individualized instruction than in regular classroom instruction include recording sets with earphones, overhead projectors, films, film strips, controlled readers, and tachistoscopes. Nonmechanical teaching aids are used in even more profusion. These include word games of various kinds, flash cards, reading series, and encoding-decoding materials. In addition most programs use material made in class by the teacher or the students.

[1] See Kiesling [25, p. 34].

PROGRAM MANAGEMENT AND COORDINATION (OR TEAMWORK)

It is extremely difficult in a small budget study to get good ideas of the quality of program management. I attempted to study program management indirectly by measuring program coordination or teamwork.

There are several benefits of teamwork. It makes possible the mutual reinforcement of goals through the dovetailing of instruction. It allows greater specialization. It encourages program personnel to share information about the problems and traits of individual children. Finally, it raises program morale. If the classroom teacher has no idea of what the specialist is doing, and no effort is being made to tell her, she may become somewhat suspicious and hostile or at least indifferent. This attitude is quickly appreciated by the program children, and instructional effectiveness is harmed. If it is obvious to the pupil that his teachers are working together, each with respect for the contribution of the other, he can respond to both without confusion.[1]

It is possible to use teamwork effectively in both group and individualized instruction, but the form that the teamwork takes is somewhat different. In group instruction, specialization is limited mostly to areas of subject matter. Two instructors can engage in dialogue before the class, for example, or one instructor can cover material within his specialty one week, another the next. In individualized instruction, specialization and teamwork can be introduced into stages of the instruction process also. One person can diagnose the child's capabilities, another can give instruction, a third can supervise and

[1]The individualized instruction that a pupil receives as part of the program is likely to be a pleasant experience, because he feels that someone cares enough to get to know him personally and to be his friend. If he feels that his regular classroom teacher is highly sympathetic to his compensatory instruction he may relate his pleasant experience to his regular school program, resulting in a much improved attitude to all of his school work.

counsel the primary instructor, and still another can evaluate the child's performance.[1]

The only program design in which it is possible to bypass most requirements for teamwork (and therefore management) is one that utilizes a highly trained and experienced specialist outside the regular classroom. She provides expert diagnosis, prescription, and instruction. She can supervise any paraprofessional aid without help. And finally, she provides all of the ongoing evaluation and would only need a good clerk to tabulate the end of the year evaluation as well. Even so, considerable teamwork is still useful in this kind of program. The specialist will often need additional diagnostic help from a psychologist or counselor. Outside evaluation is always helpful. It is almost always useful to inform both the principal and the child's regular teacher about the child's progress and needs, any special situations that require attention, and so forth. Thus although it is possible to bypass a well coordinated effort with this type of program, there might be a very real cost in terms of effectiveness in doing so.

Other program types require more teamwork. A program where the initial instruction is done by paraprofessionals in the regular classroom, for example, will require a specialist or a psychologist for diagnosis-prescription, a specialist to supervise aides, and much in-service training for aides and regular classroom teachers. A separate evaluator may be required as well as a full-time person as manager and coordinator, whose talents are of course crucial to program success. If carefully designed, this type of program may be much less expensive than the "pure specialist" treatment described above, however.

There are organizational aspects to teamwork as well. Examination of formal and informal lines of authority in these programs would seem to be a most fruitful area for further research.[2] Questions to be

---

[1] Some of the instruction can be performed separately in group instruction, too. Separate people can supervise and evaluate, for example. This is seldom done in practice, however.

[2] Some work along these lines has been done. See, for example, Halpin [13], or Katz and Kahn [22].

explored would include whether the program manager has effective control over everyone in the program and whether he makes certain that the efforts of the various instructors with whom the program children come in contact are well coordinated.

Finally, there is room for teamwork in the evaluation phase of the program. With good individualized instruction day-to-day evaluation of the child's program is almost automatic and may be done by the specialist working alone. But from the standpoint of broad policy objectives, good overall program evaluation may then be lacking.[1]

In an earlier telephone interview study, I was struck by the near-unanimity of respondents who, being asked which aspect of their program they deemed most essential, answered "good in-service teacher training." In-service teacher training was mentioned in the Hawkridge conclusions somewhat less often, although a careful re-reading of a set of their key projects revealed that indeed the concept was present in virtually all of the successful programs and either specifically mentioned as absent or not mentioned at all in most of the unsuccessful programs. These findings suggest that in-service training is quite important.

IN-SERVICE TRAINING

In-service training probably has a differential effect upon instructional personnel according to their background. For example, paraprofessionals may receive a considerable amount of in-service training but may nevertheless fail to provide instruction of the caliber of that provided by trained reading specialists (who presumably need much less in-service teacher training).

---

[1] For a good discussion of teamwork in the evaluation phase, see Rapp [28].

IV. DATA COLLECTION

THE SAMPLE

In the 1969-1970 school year there were approximately 125,000 children in over 700 California Title I projects. This study is based upon a sample representing about 6 percent of these projects and 10 percent of the pupils.[1]

To insure comparability, only those projects that used the Stanford Reading Test were chosen. With this restriction, the sample was chosen on a stratified random basis, according to percentage of school pupils on AFDC (Aid to Families with Dependent Children), percentage black, and percentage with Spanish surnames. The sample is reasonably representative of the state in terms of pupil distribution, although blacks are somewhat overrepresented and Anglos underrepresented in terms of projects.[2] The final sample includes 42 schools in 37 school districts all over California. There was a slight overrepresentation of schools in Los Angeles and Orange Counties and underrepresentation of schools in extreme northern and eastern California for reasons of travel convenience. All but two of the interviews were given in person (otherwise on the telephone), and each interview took from 45 to 60 minutes.

There are two possible sources of bias in the sample. One is the limitation to the Stanford Reading Test. Although the Stanford was mandated by the State of California to be used in grades 2, 3, and 6 in 1969-1970, only about 35 percent of the Title I projects used it. It is widely though to be a "difficult" test and perhaps districts that use it have more than average self-confidence, which is in turn based on actual high quality. On the other hand, the districts that used the test may be those that are efficient enough to use the same test for two chores or perhaps not ambitious enough to adopt what is considered a more responsive test for the compensatory education program.

---

[1] Note that two schools in the same school district are considered to be two projects.

[2] This is because a disproportionate number of blacks were in a few large schools.

Another potential source of bias comes because only those projects that had readable reports were picked. (Every year about 15 percent of all projects turn in reports that are too poorly written to allow meaningful interpretation.) If poor reports are the product of poor programs, there is obvious bias.

THE QUESTIONNAIRE

The questionnaire was based directly upon the framework for studying the compensatory education process described above. Respondents were asked to report information on percentage minority and AFDC (these items could also be cross-checked from state sources), on instruction type, what aids were used, which personnel took part in instruction, size and length of classes, and class location. These data were double-checked since respondents were also asked to give schedules for the entire day of instruction personnel. Questions were designed to show who conducted diagnosis-prescription, to whom prescriptions were communicated, which kinds of tests were used, and length of testing time. Similar questions were asked with respect to planning and in-service training. Finally, questions were asked concerning lines of authority, including who decided and who closely helped decide on issues concerning hiring of program personnel, choosing program children, and a number of other program characteristics.

The questionnaire was pre-tested twice with analysis of problems and revision occurring after each pre-test. It was designed to be given in person and to require only the responses of the operating manager of the school district Title I program if that person was well informed. In large school districts, however, it was necessary to interview both the building program manager and the district program manager. In many instances information was obtained from others besides the primary respondent.[1] The questionnaire is reproduced in Appendix A.

---

[1]Often as I conducted my interview and came to a section of questions the respondent did not feel competent to answer, he or she would get me a quick appointment with someone who knew the answers (or at least give me their name and telephone number for a telephone query later) or

often pick up the telephone and call someone to find out while I waited. An advantage of giving the questionnaire in person is that it is quickly ascertained to the mutual agreement of both interviewer and interviewee when the latter is weak with respect to knowledge of some program aspects.

V. VARIABLE CONSTRUCTION

THE PERFORMANCE MEASURE

California compensatory education projects are required to submit performance data once yearly to the Division of Compensatory Education including information concerning program objectives, instruments used, number of project participants by grade, project length, and frequency distributions of scores at the beginning and end of the treatment period. They are also asked to provide median pre- and post-test scores and the gain in grade equivalent by grade.

As mentioned above, some 35 percent of all the projects that submitted reports to the state used the Stanford Reading Test. It was thus possible to use the gains in standard grade scores on the Stanford test for the performance measure. Since the reports also include information concerning the specific objectives of these programs, it was possible to choose the sample only from schools that put as their major objective the raising of reading scores on standardized reading tests. To some extent therefore, one of the comparability problems noted in the literature -- studying programs with different objectives[1] -- was overcome.

Two performance measures were used, ending score and gain in score per month of program duration (both in grade equivalents). The latter measure was used as an effort to consider separately from program length the possibility that learning does not occur evenly over the length of the program, and the former measure was used because gain scores have been criticized in the educational psychology literature. The measures were used for pupils pooled over grades 2, 3, 4, and 5, and for grade 3 alone, as that grade was the only one in which there were enough observations for meaningful analysis. The justification for these procedures and the discussion of some other relatively minor problems concerning the performance measure are given in Appendix B.

It is conceivable that performance gain on standardized tests is not only a function of program treatment but also of where the children

[1] See McDill et al. [26].

started. Often this relationship is positive: the pupils who start higher gain more.[1] If there is a test ceiling or "topping out" effect at work, however, the relationship might well be negative. In either case, proper specification of the model demands that the variable be included. As used in the estimating equations, the variable was coded as the number of months the children were below the national norm at the beginning of the program plus 20.0.

SOCIOECONOMIC VARIABLES

It is desirable to account for systematic differences in socioeconomic characteristics of pupil environments in order to assess the impact of the school program properly. Attempts were made to control for socioeconomic differences among pupils in two ways, First, respondents were asked to characterize the educational and occupational levels of the parents of their program children. This was, for several reasons, unsuccessful.[2] Second, a considerable amount of factual socioeconomic information was collected. Such data included the percentage of children in the school attendance area who were receiving AFDC and the percentage of program children belonging to minorities.

Another characteristic that must be admitted to the analysis is the degree of mobility of program children. This may be a proxy for socioeconomic characteristics since there are studies that show mobility to be positively related to low socioeconomic status [5]. Mobility

---

[1] In an earlier study I found that gains in performance from grade 4 to grade 6 were highly correlated with score in grade 4. [24].

[2] Data concerning family characteristics that might bear upon pupil motivation are simply not collected. The reason for this is understandable. Many children in Title I programs come from homes that unfortunately have characteristics about which they feel embarrassed. Many program instructors feel that merely asking children questions concerning their home environment causes an adverse effect upon pupil morale and achievements. It should be possible to overcome this problem by administering instruments or questions to the children that might, directly or indirectly, assess such characteristics as amount of verbalization in the home, and so forth, without directly embarrassing the child if there is some problem. The use of one such test is described in [6].

itself can be injurious to program quality, of course.[1] Thus, even though a particular child stayed in the program all year, the quality of his instruction could be affected by the fact that his teachers are constantly bothered by the comings and goings of other children in the program.

INSTRUCTIONAL INTENSITY BY TYPE OF INSTRUCTOR

As has been discussed already, the amount of instruction on an individual equivalent basis was central to the analysis in this study. The interviews recorded how the pupils spent their project time, and this information was used to fashion the variables of individual equivalent minutes (IEMs) spent with each child on a weekly basis by instructional personnel.

The variable allows for one measure to be constructed out of size of class, number of instructors, and length of session. Some allowance was made also for supervision time when the specialist or classroom teacher used one or more paraprofessional persons as assistants in actual instruction.

An example of how the variable is constructed is as follows: If a single specialist sees groups of 10 pupils 30 minutes per day 5 days per week, IEMs would be 15 (30 divided by 10 times 5). If the specialist has one paraprofessional assistant for these 10 pupils IEMs for each pupil, abstracting from supervision time, doubles. Since it is assumed that the specialist and the paraprofessional both lose 10 percent of their time in the specialist's supervision of the paraprofessional, IEM for each is not 15, but 13.5.[2]

---

[1] Mobility does not directly affect the performance outcomes since tests scores were reported by the projects only for pupils present both at the beginning and the end of the program. The question that was asked to obtain mobility rate was: "What percentage of those children who were initially placed in the program at the beginning of the program year were still in the program at the end of the program year?"

[2] The convention used was to deduct 10 percent of the instructional time of supervising teacher and paraprofessional for each of the first two paraprofessional aides, and 5 percent of each aide after that.

There are three types of personnel used in instruction in the program: the trained reading specialist, the regular classroom teacher, and the paraprofessional. However, for constructing variables, paraprofessionals were divided into those assisting regular classroom teachers and those assisting reading specialists.

PERCENTAGE OF INSTRUCTION IN THE REGULAR CLASSROOM

Considerable importance attaches to the relative effectiveness of supplementary instruction in the regular classroom as opposed to that given in a separate facility. If effective instruction could be given in the regular classroom, the cost would be much less and the regular classroom teacher could assume a more active part. She could also receive valuable in-service training in the course of her regular duties. On the other hand, a specialist can give more undivided attention to children in a separate facility. We would expect to find a positive relationship between use of separate facilities and pupil performance, although this difference would probably be lessened in projects that have considerable teamwork and in-service training of regular teachers. The actual percentage of instruction given in the regular classroom was the variable used.

USE OF EDUCATIONAL MATERIALS AND EQUIPMENT

The possible importance of different types of educational materials and equipment was mentioned above. In the study, however, it was impossible in practice to differentiate between the amounts of materials and equipment used. Thus it was found that the essential characteristics of the lists of materials and equipment obtained for each program were virtually identical (at least to the untrained eye). There were some differences in the amounts used to be sure, but these were merely that there were more such materials in separate facilities and that reading specialists tend to use them more than regular classroom teachers. Because of this virtually complete overlap between percentage of instruction in the regular classroom and percentage of instruction given by the trained specialists, I decided not to include a variable in the

model for type of educational equipment used. However, any positive findings for percentage of instruction in the separate facility and instruction given by trained reading specialists must necessarily include in part a finding that there is possibly some return to the heavier use of such materials and equipment.

COORDINATION AND LEADERSHIP VARIABLES

Several variables were used to represent program coordination. The simplest of these was hours spent in program planning per week. In the interviews, the respondents were informed what was meant by planning and by in-service training and then were asked how much of each took place. Since planning and in-service training are often difficult to separate, and also because there are problems with respondent's collective memories and with quantifying the length of informal discussions, both variables are probably subject to considerable measurement error.[1]

A variable was also used to account for presumed weaknesses in lines of authority within the projects. Teamwork should depend in part upon the degree to which all the principal actors in the project are subject to control by the same person. (Also, of course, it should depend on whether he or she uses the control wisely.) The questionnaire was designed to discover not only the formal but more important the informal "chain of command." On the basis of the information collected, a dummy variable was constructed. It was set equal

---

[1] As was explained to the respondents, planning was defined to include the kinds of topics and skills program personnel should be covering during the coming week or weeks for individual children (by name). In-service training meant explanations concerning why project personnel should take various educational steps, how and when a certain skill requires that another kind of skill be taught immediately prior, and so forth. Demonstrations concerning classroom techniques on how to teach skills that the program leaders desire to be taught are also included.

to unity when conflicts in direction and purpose were reasonably possible, and zero otherwise.[1]

One additional coordination variable was defined. Respondents were asked to identify the personnel who attended planning meetings. It was hypothesized that a well coordinated program would routinely have more "key" personnel present at such meetings. The percentage of attendees who were considered "key" people became the variable.

USE OF PSYCHOLOGISTS FOR DIAGNOSIS

There was considerable variation in the amount of psychologist time used in the diagnosis and prescription phases of the programs. To test the hypothesis that intensive use of psychologists' diagnoses may be associated with better performance, a dummy variable was constructed on the basis of number of pupils per full time equivalent psychologist.[2]

---

[1] An example of the "no conflict" situation would be where the program is directed by an Assistant Superintendent with line authority who is not too busy to devote a reasonable amount of time to the program. Thus, no coordination problem need ever arise: all personnel concerned, including specialists, building principal, and so forth, are directly responsible to the Assistant Superintendent.

A majority of the actual programs were included in the "conflict possible" category, however. Often, for example, the program director has a rank equal to the building principal and has no "line" authority. The Director might supervise the specialist within a given school, while the building principal supervises the classroom teacher and paraprofessionals. The success of such a program depends crucially upon how closely the director and the building principal cooperate. Even if these two individuals are good friends, chances are that the effect of the specialist and regular classroom teacher may not be well coordinated, or at least this is my supposition. A variation of this pattern exists when a person has the control but has too many other duties to use it effectively to coordinate the program.

[2] There were very few projects with a ratio of pupils to full-time equivalent psychologists near 1000:1. Since most projects fell either clearly above or below this figure, if the ratio was below 1000:1 the dummy variable was set equal to unity and if above, to zero.

VI. FINDINGS

The model of school performance with the best explanatory power is presented in equation (1). All other variables discussed failed to add explanatory power to the model.

$$
\begin{aligned}
(1)\quad \text{SCORE 25} = {} & 3.45 + \underset{(3.3)}{4.85}\ \text{PGMLENGTH}^{*} + \underset{(7.4)}{.86}\ \text{BEGIN 25} \\
& \underset{(1.0)}{-.013}\ \text{PCTMIN} + \underset{(3.1)}{1.30}\ \text{SPECIEMS}^{*} \underset{(1.7)}{-.023}\ \text{PCTREGCR} \\
& + \underset{(2.3)}{.106}\ \text{TCHRPPIEMS} + \underset{(2.5)}{2.07}\ \text{PLANHRS} \\
& \text{SE Estimate} = 1.84 \\
& F(7,34) = 21.32 \\
& \text{Corrected } R^2 = .78
\end{aligned}
$$

All of these models are weighted to correct for heteroscedastic error terms due to unequal numbers of pupils in each project.[1] The values given in parentheses are t statistics,[2] and variables marked with an asterisk are transformed into their logarithms. Variable descriptions are given in Table 2.

Instruction by both specialists and paraprofessionals assisting classroom teachers is related to pupil performance. For the paraprofessionals ten individual equivalent minutes of instruction weekly are related to an additional month of reading performance. Specialist instruction shows a declining relationship with ten IEMs related to about 1.5 months of reading gain for the first ten minutes of instruction and then declining to less than one-third month of gain per ten IEMs beyond approximately 40 IEMs. The specialist variable was somewhat more statistically significant as well.

There is a small gain in performance when programs are conducted outside the regular classroom, although this variable is only barely significant at the 10 percent level.

---

[1] Weighting is further discussed in Appendix B.

[2] For 34 degrees of freedom, significance levels are: 5 percent, 2.0; 1 percent, 2.7; .1 percent 3.5.

Table 2

MEANS, STANDARD DEVIATIONS, AND DESCRIPTION OF VARIABLES

Variable Name	Mean	Standard Deviation	Description
SCORE 25	17.46	3.36	Score at the end of program for students in grades 2, 3, 4, 5, in number of months relative to the grade level norm, coded such that the end score norm was 28.4 and the begin score norm was 20.0.
SCORE 3	17.79	3.22	Score at the end of program for students in grade 3, in number of months relative to the grade level norm, coded such that the end score norm was 27.8 and the begin score norm was 20.0.
GAINSCORE 25	0.87	0.40	Months gain on Stanford Reading Test per month of instruction, weighted average, students in grades 2, 3, 4, and 5.
GAINSCORE 3	0.84	0.56	Months gain on Stanford Reading Test per month of instruction, students in grade 3.
PGM LENGTH	8.43	1.65	Length of program in months, from pre-test to post-test.
BEGIN 25	10.88	3.25	Months behind national norm of students at beginning of program, grades 2, 3, 4, and 5, plus 20.0.
BEGIN 3	10.37	2.59	Months behind national norm of students at beginning of program, grade 3, plus 20.0.
PCTMIN	59.1	27.7	Percent of program children American indian, black, and Spanish surname.
SPECIEMS	18.0	13.7	Number of individual equivalent minutes (IEMs)[a] per week taught by trained reading specialists.
TCHRIEMS	16.3	10.1	Number of IEMs per week taught by regular classroom teachers.
TCHRPPIEMS	8.8	8.4	Number of IEMs per week taught by paid paraprofessionals assisting regular classroom teachers.

Table 2, continued

Variable Name	Mean	Standard Deviation	Description
PCTREGCR	54.6	34.7	Percentage of Title I instruction given in the regular classroom.
PLANHRS	0.57	0.38	Hours per week project personnel spent in planning meetings.

[a]See page 19 for a description of individual equivalent minutes.

The only coordination-management variable related to performance was number of planning hours, with one hour per week of planning (which is more than most projects had) being associated with an additional 2.1 months gain. Causation cannot necessarily be inferred from the relationship, but it does suggest that some formal planning does indeed pay dividends. It is interesting to note that the in-service training variable, about which there were high hopes built on analysis of prior findings, always had the wrong sign and was never significant.

According to the variables both included and omitted from equation (1), no SES variable is important. Of the variables not included, percentage of children with Spanish surnames had no explanatory power, while percentage black was weakly and insignificantly negatively related to performance. The percentage of children who moved, which can be considered as a proxy for one SES characteristic, was negative and usually yielded coefficients larger than their standard errors. The variable for percentage of children in the school attendance area on AFDC, which had been considered one of the more meaningful SES variables, consistently displayed the wrong sign, although it also was not statistically significant.

The percentage minority variable was somewhat collinear with amount of instruction conducted in the regular classroom (R = .50) and was somewhat more significant when that variable was not included in the model. To show this difference, equation (2) is a slightly different specification, with percentage of instruction inside the regular classroom being replaced by instruction by the regular classroom teacher.

$$
\begin{aligned}
(2)\quad \text{SCORE 25} = {} & -4.89 + \underset{(1.5)}{4.47}\ \text{PGMLENGTH}^{*} + \underset{(3.0)}{.85}\ \text{BEGIN 25} - \underset{(7.0)}{.023}\ \text{PCTMIN} \\
& + \underset{(1.9)}{1.59}\ \text{SPECIEMS}^{*} - \underset{(3.9)}{.033}\ \text{TCHRIEMS} + \underset{(0.6)}{.090}\ \text{TCHRPPIEMS} \\
& + \underset{(1.4)}{1.58}\ \text{PLANHRS} \\
& \qquad (1.9) \\
& \text{SE Estimate} = 1.91 \\
& F(7,34) = 19.53 \\
& \text{Corrected } R^2 = .76
\end{aligned}
$$

In this model the percent minority variable is significant at almost the 5 percent level. Specialist instruction becomes even more significant than before, but instruction by paraprofessionals helping classroom teachers loses some of its significance. Since more effective individualized instruction (including use of more specialized materials and equipment) is carried on in the separate facility, the first model represented by equation (1) is undoubtedly much preferable to that in equation (2) on a priori grounds.

Programs depending almost exclusively upon reading specialists for their instruction might be expected to require less management and teamwork. To test this, the model was fitted to 25 projects that did not depend heavily upon specialist instruction.[1] The results are shown in equation (3).

$$\begin{aligned}(3)\quad \text{SCORE 25} = {} & -7.65 + \underset{(1.3)}{5.33}\ \text{PGMLENGTH}^{*} + \underset{(2.0)}{.81}\ \text{BEGIN 25} - \underset{(6.1)}{.011}\ \text{PCTMIN} \\ & \underset{(0.7)}{} \\ & + \underset{(2.7)}{1.66}\ \text{SPECIEMS}^{*} - \underset{(0.3)}{.0063}\ \text{PCTREGCR} + \underset{(2.1)}{0.109}\ \text{TCHRPPIEMS} \\ & + \underset{(1.4)}{1.86}\ \text{PLANHRS} \\ & \text{SE Estmate} = 1.89 \\ & F(7,17) = 13.89 \\ & \text{Corrected } R^2 = .79\end{aligned}$$

The importance of the planning hours variable is somewhat lessened instead of vice versa, and indeed this was true for all the other coordination and leadership variables as well. The hypothesis of better coordination in nonspecialist dominated programs fails to be confirmed by the data.

Finally, because of the problems mentioned above with respect to aggregating data from different grade levels, the model was fitted to the 38 projects for which data were available for grade 3. The resulting equation, presented as equation (4), only manages to replicate the

[1] The criterion used in making the distinction was that more than half of total instruction was accomplished by specialists together with paraprofessionals assisting specialists, and at the same time more than half of all instruction took place in a separate facility.

finding for the importance of specialist instruction, with the earlier significance of instruction of paraprofessionals helping classroom teachers and planning hours reduced to insignificance. This finding introduced a note of caution into the interpretation of the meaningfulness of the latter two variables, therefore. It is interesting that the t value of the beginning score variable increases greatly and changes sign while that for program length is reduced to insignificance.[1]

(4) SCORE 3 = 5.28 + .53 PGMLENGTH* + .78 BEGIN 3 - .0060 PCTMIN
(1.0) (0.2) (3.9) (0.3)

+ 1.60 SPECIEMS* - .081 PCTREGCR + .048 TCHRPPIEMS
(2.6) (0.9) (0.7)

+ .76 PLANHRS
(0.6)

SE Estimate = 2.59

F(7,30) = 4.08

Corrected R^2 = .37

DESCRIPTION OF THE SIX BEST PROJECTS

The top performing six projects in the study had pupil gains of at least 1.25 months per month of instruction. They averaged 1.5 months gain per month of instruction. Following is a brief outline of the characteristics of these six projects.

Although four of the six projects had large amounts of instructional time for each pupil per week, the intensity of instruction in the other two was below average. It would appear therefore that large amounts of instruction are not absolutely necessary for good performance, but they are quite helpful.[2]

---

[1]The PGMLENGTH and BEGIN variables are collinear (r = .56), and some of this strange behavior could be caused by that fact.

[2]The average number of IEMs for all 42 projects was 44, and the two projects mentioned as below average had 37 and 25 IEMs respectively. The difference in instructional intensity between the best and worst projects is striking, however. The average number of IEMs for the six best projects, including the two just mentioned, was 70. The average for the ten worst projects, which had an average gain of about .4 months

In five programs a large proportion of the instruction was given by trained reading specialists. In the sixth, a paraprofessional who had three years' training by a specialist gave individualized instruction in a separate facility.

In the four projects in which the specialists employed paraprofessional aides, the amount of instruction given by the aide varied between one-fourth and one-third of the amount given by the specialist. In all projects the specialists gave instruction in small groups no larger than ten students. Only two projects used classroom teachers and paraprofessionals in assistance of classroom teachers, and these two projects had large doses of specialist instruction besides. Four of the six programs had all instruction in a separate facility; the other two had half of their instruction in a separate facility.

There was no discernible trend among the six projects with respect to minorities represented. Three of the projects had a very high proportion of the students belonging to minority groups and in the other three the percentage was quite small. Two projects had high percentages of black students and four had no blacks. Two projects had a high percentage of Spanish surname children. There was also considerable variation in pupil mobility in the six projects.

Concerning some other school variables, the number of pupils per full-time program manager in all six projects was quite low. On the other hand, the number of pupils per psychologist in the projects varied widely. The number of planning hours per week and the number of hours of in-service training per week also varied quite widely. In all six projects almost all key people were present at all the planning meetings.[1] In several projects, the chain of authority looked

---

per month of instruction, was only 32. The difference in the amount of instruction given by trained specialists is even more striking: 30 IEMs in the best projects as opposed to 12 in the worst.

[1]This was not true in the ten worst projects where the percent of key people average was 75. It is notable that in these ten projects, when the percentage of key people present was high, the actual planning time was small.

to be somewhat muddled, and therefore this variable does not seem to be very representative of high quality programs.

In terms of geographical setting, the projects were all medium or small in size and were all either in rural or suburban settings. There were no large urban schools represented in the six top schools in the study.

To summarize the characteristics found in all of these highly successful projects, all six had small group instruction by specialists, high ratios of managers per pupil, and a consistently large percentage of key people present at planning meetings.

DISCUSSION OF FINDINGS

There has been wide commentary in the educational literature that compensatory education has failed, that there is no evidence to show that anything done in compensatory education programs is related to the performance of children from disadvantaged backgrounds.[1] The findings here with respect to the relationship of instruction by trained specialists to pupil performance, which maintains significance no matter which of the meaningful subpopulations of these programs is chosen for fitting the model, clearly contradict this widely repeated set of findings. Instead, the evidence here supports the "reasonable hunch" of Guszak based on work by Turner and others that the instructional procedures used by the diagnostic reading specialist are important. The evidence also suggests that instruction given by paraprofessionals helping regular classroom teachers may be effective.

School personnel who deal with disadvantaged populations often use 0.7 months per month of instruction as the "normal" rate of advance for

---

[1] To cite only two: "Compensatory education has been tried and it apparently has failed." Jensen [20, p. 2]. "Negative residual gain-scores for most 'participating' groups in all grades seem to indicate that even when a lower 'starting point' is considered, participants did not progress at the same rate as nonparticipants." Glass et al. [10, Chapter 6, p. 148].

these children using traditional instructional methods.[1] The average gain in these projects was 0.87 months per month of gain. If the 0.7 figure is correct, the overall impact of the Title I money would be .17 months gain per month of instruction. For the projects that make heavy use of specialists giving individualized instruction, however, the gain is more. Increasing specialist instruction per child by 20 minutes per week should raise the average by at least .2 months, to a rate at which pupils would be slowly catching up. It would be dangerous to extrapolate the findings too closely in this way, but there is room for optimism.

Findings for the remaining aspects of the study are not nearly so positive, however. Although the planning variable is significantly related to pupil performance in the main explanatory model used, the finding fails to hold up when the model is fitted to other meaningful subpopulations; also, none of the other variables constructed to measure aspects of coordination and management were related to pupil performance at any time. With the possible exception of the finding for planning time, then, the general conclusion will have to stand that the strong hypotheses carried into the study with respect to the importance of coordination, teamwork, and management to program success failed to be supported by the regression analysis. The descriptive results were somewhat more positive with respect to the importance of the amount of management input and to the percentage of key people who participated in planning sessions.

Whether the coordination variables failed because they represent reality, or because the variables are themselves too poor, remains to be seen in further studies. The latter possibility is considered highly likely, although the very negative relationships found for some of the variables lead one to suspect strongly that the negative findings to

---

[1] The figure found in the Coleman Report was disadvantaged children who reach grade 12 are about 3 grade levels behind. This would imply a figure of .75 months per month of instruction for those <u>who do not drop out</u>.

some extent represent reality as well.[1] This is indeed increased by the fact that nonspecialist dominated programs had even more negative values for these variables in all cases than when the model was fitted to all projects. The same was also true for the in-service training variable, and the consistent null finding for that variable was something of a surprise and disappointment in the study considering all the rhetoric in the past two years directly and indirectly from program managers concerning the importance of good in-service training. Perhaps the problem was that we were not able to discriminate between good in-service training and poor in-service training, or perhaps it is in part because specialists (who are most effective in securing good results) do not require as much in-service training as other instructional personnel.

Proper discussions of the findings for program length and beginning score fall outside my professional competence. Program length is related to performance, and the evidence suggests that more learning is done early in the program than later since the variable fit the data much better when transformed into its logarithm. (This is suggested by the negative coefficient for PGMLENGTH in equation (1B) in Appendix B also.)

It is unfortunate that the model, when fitted to the grade 3 scores, did not replicate the findings for the teacher paraprofessional and planning variables obtained in equation (1). In interpreting this difference, how likely is it that the aggregation of data over different grade levels will lead to error? The question is discussed in more detail in Appendix B. I feel that the performance levels shown by the pooled grade data represent reality more faithfully than those for grade 3, but some readers may disagree after reading Appendix B.

If the pooled data findings are most representative of reality, the findings in the study are not all in one direction. Instruction by

---

[1] A cynical explanation, which I would be inclined to reject, is that all projects had uniformly bad management so there was nothing good to measure. I would also be inclined to reject the opposite explanation that all projects had management that was uniformly good.

the classroom teacher with her paraprofessional (with that given by the paraprofessional doing most of the counting in this case) does in fact seem to be related to performance, to a degree about two-thirds as great as that for the trained specialist. If the significance level for the paraprofessional variable were the same, we could immediately draw some rather direct economic conclusions from this, of course, but since the confidence with which we can accept the paraprofessional finding is lower, extrapolation would be dangerous.

Finally, the difference in the relationship of socioeconomic status variables to performance in this study as compared with other input-output studies should be noted. Although most other studies have SES as the quality most highly related to performance, no SES variable was significant here. Part of this can probably be explained by the fact that the other studies had pupil populations with wider variation in SES. This is even true when, as in studies by Bowles [4] and Hanushek [14, 15], populations were restricted by race, since there were of course middle and high SES black or Spanish surnamed children present in their samples. The present input-output study is the only one that exclusively used low status children. On the other hand, the variables used may have been inadequate. Even the percentage of children in the school area on AFDC, upon which substantial hopes had been riding, completely failed to be related to performance. Much more sophisticated SES measures may be necessary for discriminating such things as verbalization in the home (see, for example, [5]), motivation, and the like. Yet, as indicated above, a procedure that depends on asking the child a straightforward question about these things is completely unacceptable for pedagogical reasons. It is perhaps surprising that the model explained as much of the variation in performance as it did, given the inadequacy of the SES variables.

VII. CONCLUDING COMMENTS

This study is the first to attempt to assess compensatory education projects with input-output methodology. A single performance measure is used across all projects, and an attempt is made to account for socioeconomic differences using multiregression techniques. As with other input-output studies,[1] the largest failure of this one is that the analysis is not student-specific, or even classroom-specific. An attempt was made to do some things not previously done in input-output studies, however, in that program organizational characteristics and instructional organizational strategies are related to pupil performance.

Since I lacked the necessary expertise to study the internal workings of the instruction, and also the necessary budget for doing highly refined techniques with organizational relationships, the study is only a first step and no more is claimed for it. I had hoped that this procedure might permit a first, rather fuzzy look at the insides of what has been termed the enigmatic "black box" of the inner workings of schools from the standpoint of input-output methodology, but only with respect to broad organizational patterns and not in a truly student-specific way. If this kind of methodology is to be pursued farther, of course, that will have to be added next.

It is certainly important for the cost-effectiveness of the nation's educational research that wise heads carefully consider the payoffs to future research of the type undertaken here. It is by no means a unanimous opinion that such research will yield results worth their cost in the future. Thus, Alcaly, in commenting on the Hanushek study mentioned above, claimed that further studies of the same genre would probably not repay the cost [1]. In commenting on an earlier version of the present study, Ribich came to much the same conclusion [29]. On the other hand, Weisbrod said that there were probably increasing returns for many more research efforts of this kind [32].

---

[1] Except Hanushek's [15], which was classroom-specific.

If the approach does seem viable, the findings suggest several avenues for future work. The one most immediately suggested is to expand the analysis of differences in instructional techniques and to include student-specific analysis. Individual students must be matched to individual teachers and treatments in large enough samples and with enough control for socioeconomic differences that findings would be statistically believable. Second, much more careful thought will have to be given to program organization, coordination, and management. Some progress has been made in the past using role-analysis techniques in education, but further exploration is needed. Specialists familiar with organizational characteristics of large organizations, whether public or private, should be brought in to work on these questions. Finally, much more sophisticated work will have to be done to find meaningful socioeconomic variables.

Appendix A

COMPENSATORY EDUCATION PROGRAM QUESTIONNAIRE

Name of school district _____

Name of school _____

Name of respondent _____

Title of respondent _____

Background experience of respondent _____

I. GENERAL

Total No. of elementary pupils in district _____

Total No. Title I designated pupils in district _____

Total No. Title I designated pupils in school _____

Number of elementary schools in district _____

Number of elementary schools in program _____

Are programs different, building to building? Yes____ No ____

Do you have evaluation results, building to building? Yes ____ No ____

Percent of pupils in the program this year which also received treatment last year _____% ....two years ago _____%

Length of school year _____ days; Program year _____days.

Answer with respect to school named above.

About the Program Children

Briefly, how chosen? _____

_____

_____

Would you characterize as best you can the backgrounds of the children according to the following:

Occupation of principal breadwinner

Unskilled	_____%
Semi-skilled	_____%
Skilled	_____%
Above-skilled	_____%

Education of principal breadwinner

0 - 7 years	_____%
8 years	_____%
9 - 11 years	_____%
12 years	_____%
more than 12 years	_____%

Racial Composition

Mexican-American	_____%
Black	_____%
Other white	_____%

2. DIAGNOSIS - PRESCRIPTION

In all compensatory education programs there is diagnosis of the problems that require "compensating" educational effort. This can be done by the classroom teacher in the course of her instructional day, or by special diagnosticians. Please supply the following.

Diagnosis Personnel

Which of the following devote time to diagnosing pupil learning difficulties?

	Number	Time Per Week (%)	For Which Weeks?
Program Director	_____	_____	_____
Building Principal	_____	_____	_____
Psychologist	_____	_____	_____
Reading/Math Specialist	_____	_____	_____
Counselor	_____	_____	_____
Classroom teacher	_____	_____	_____
Para-professional	_____	_____	_____
Others:			
_____	_____	_____	_____
_____	_____	_____	_____
_____	_____	_____	_____

Name of objective test used for diagnosis, if any. (Not the same test as used for evaluation.) _____

Testing time per pupil _____

Individual interviews used? Yes____ No ____

Conducted by whom? _____

Time spent per pupil individual interviews _____

Physical examination given? Yes____ No____ Length _____

How initiated?_____ Routine for all? _____

Referral? _____ Other _____

Any other special diagnostic techniques used? Yes____ No ____

What? _____

For what percent of the program children? _____

Diagnosis always presupposed accompanying prescription of method for dealing with the individual learning situation found.

Which of the personnel listed above has final operating authority for determining the prescription for each child? _____

Which personnel helped determine the prescription? _____

_____

In the course of the program, list which teaching and management personnel had individual pupil prescriptions communicated to them:

Routinely _____

Regularly, but infrequently _____

_____

Occasionally _____

_____

_____

_____

3. INSTRUCTION (In the Representative School)

List all personnel who did actual instruction of children in your Title I program, with years of experience in this kind of assignment.

Number	Type of Instructor	Years of Experience (List or give average if more than one in category)
_____	Trained reading or mathematics specialist	_____
_____	Regular classroom teachers	_____
_____	Paid para-professional aides	_____
_____	Unpaid para-professional aides	_____
_____	Peer-group tutors	_____
_____	_____	_____
_____	_____	_____
_____	_____	_____

Description of Instructional Program (Instructional Units Summary Page)

Size of Instructional Units: Description and Example

Give size of instructional units for program children, indicate total program time spent in each, and indicate whether and which other instructional personnel were helping the major instructor in the classroom. (By "instructional unit" we mean the size of group of children sharing the same instruction. For example, assume program children met in groups of 12 five hours per week with a specialist and two paraprofessionals. Assume the specialist and paraprofessionals teach as a team for 30 minutes and then split up into three groups of 1 instructor with four children. The question in this case is answered as follows.)

Example

	Size of Instruction Unit (Pupils)	Titles of Principal Instructor(s)	Time Spent (Per Day) (min)	No. & Titles of Asst. Instructors, of amy
1.	12	Reading Specialist	30	2 aides
2.	4	Specialist, 2 aides	30	--

In the appropriate columns opposite the description for each different instructional unit size, give the facility used, type and size, list typical instructional aids and the percentage time used (roughly) and audio-visual equipment and the number of times each was used weekly. (Approximate as best you readily can.)

Were there any instructional techniques used that were unique in some way? If yes, please describe. _____

_____

_____

Were field trips taken beyond those in your regular school program? Yes____ No____

How many? ____ Average Cost? _____

4. PLANNING AND IN-SERVICE TRAINING OF INSTRUCTIONAL PERSONNEL

Were there regular planning meetings: Yes____ No____

If so, who usually conducted them (Title)? _____

_____

When this person was not present, who conducted them (Title)? _____

List by title the personnel normally present at these meetings.

_____ _____

_____ _____

_____ _____

_____ _____

INSTRUCTIONAL UNITS SUMMARY

Size of Instructional Unit	Titles of Principal Instructor(s)	Time Spent Per Day by Pupil (min)	Number and Titles of Asst. Instructors, if any	Facility Type and Approximate Size	Examples of Instructional Materials Routinely Used

INSTRUCTIONAL UNITS SUMMARY (Continued)

Percentage Time Instructional Materials Used (By type if relevant)	Items of Audio-Visual Equipment Used	Rough Approximation of No. of Times Per Week Items of A-V Equipment Used	Instructional Technique Employed

On this page, give the schedules of each of these types of person for typical program days, making sure to distinguish between time spent with program and non-program children. Indicate for all five week-days. If the same schedule for all five days, put "all" under days.

Schedule	Days

Frequency of meetings and length.

Daily _____ Weekly _____ Bi-Weekly _____ Monthly _____ Other _____

Length of meetings in minutes _____

_____

_____

_____

Could you estimate roughly what percentage of these meetings were given up to in-service training for instructional personnel? _____

If there was such training, who conducted it? _____

_____

_____

Were there other meetings in your district and/or school devoted chiefly to in-service training of instructional personnel? Yes_____ No_____

If so, list the persons conducting the meetings and number of hours per month spent by each.

_____

_____

_____

List the number and types of personnel who were the attendees (trainees) at these meetings, and time per month on the average spent by each.

Number	Title	Hours Per month
_____	_____	_____
_____	_____	_____
_____	_____	_____
_____	_____	_____
_____	_____	_____

Can you give the amount of time per week typically spent in communication between leaders of in-service training and diagnosis-prescription personnel? _____

5. EVALUATION

List persons by title who conducted overall evaluation of the program, and time spent.

Person	Percent of Week	Which Weeks
_____	_____	_____
_____	_____	_____
_____	_____	_____

Dates of testings

Pre-test: _____

Post-test: _____

What percentage of program children move each year? _____

Aside from the annual report to the State, briefly describe frequency of written or oral evaluation reports given to the following people:

Title	How Often
Coordinator	_____
Building Principal	_____
Diagnosis Leader	_____
In-service training leader	_____
Instructors	_____
Parents	_____

Is there an outside evaluator? Yes____ No____

How much time in hours does he spend per year with:

Title	How Often
The Program Leader	_____
The Building Principal	_____
Diagnosis Personnel	_____
Instructors	_____
Parents	_____

How many planning and/or in-service meetings were attended by the outside evaluator: ____

Copy of evaluation report for 1960-69? Yes ____ No____

Person to contact concerning this year's report. _____

6. ENVIRONMENT

Briefly describe methods used to affect the pupils' home environments, if any.

_____

_____

_____

_____

_____

_____

7. OVERALL COORDINATION

Title of the effective leader of the District Compensatory Education Program_____

_____

Title of the effective leader of the program in _____school.

_____

Title of titular leader if different from effective leader (district) _____

_____

Who do you take direct instructions from concerning Title I matters? _____

_____

Who do you give direct instructions to and have authority over concerning Title I matters? _____

_____

Who do you give instructions to in the nature of advice that is almost always taken?

_____

Who effectively makes the final decisions and who collaborates heavily concerning:

Choosing instructional techniques	Final:	_____
	Coll:	_____
	Coll:	_____
Choosing instructors	Final:	_____
	Coll:	_____
	Coll:	_____
Purchasing Materials	Final:	_____
	Coll:	_____
	Coll:	_____
Deciding on evaluation personnel	Final:	_____
	Coll:	_____
	Coll:	_____
Choosing Program Children	Final:	_____
	Coll:	_____
	Coll:	_____
Designing In-service Teacher Training	Final:	_____
	Coll:	_____
	Coll:	_____

Give time per week (hours or minutes) _____

The effective program leader spends communication with:

Diagnosis personnel _____

In-service training staff _____

Instruction staff _____

Evaluation personnel _____

Were they any of the above (excluding building principal) over whom the effective leader did not have direct control? _____

_____

If not, who did? _____

Appendix B

STATISTICAL DISCUSSION

This appendix includes discussions of some statistical questions that were considered to be of insufficient general interest to be involved in the main text of the report.

USE OF GAIN SCORES

Two performance measures were used in the empirical work done in this study. One of these was gain in grade equivalents per month of elapsed program time. Since there has been considerable criticism in the educational psychology literature on the use of gain scores because of the regression to the mean phenomenon (see Cronbach and Furby, [8]), only end score was used in the findings presented in the text. Use of gain per unit of time elapsed does allow a direct look at the rate of learning over the length of programs, however, and besides this, a presentation of the model fitted to the gain variant should give some insight into the possible damage of using gain score. The fitted equation, which is similar to equation (1) in the text, is therefore presented here as equation (1B).

$$
\begin{aligned}
\text{(1B)}\quad \text{GAINSCORE 25} = {} & 0.85 - \underset{(3.5)}{} - \underset{(1.3)}{.031}\ \text{PGMLENGTH} - \underset{(1.0)}{.015}\ \text{BEGIN 25} - \underset{(1.1)}{.0016}\ \text{PCTMIN} \\
& + \underset{(3.3)}{.16}\ \text{SPECIEMS}^{*} - \underset{(2.0)}{.0032}\ \text{PCTREGCR} + \underset{(3.2)}{.017}\ \text{TCHRPPIEMS} \\
& + \underset{(2.6)}{.25}\ \text{PLANHRS} \\
& \text{SE Estimate} = .216 \\
& F(7,34) = 8.45 \\
& \text{Corrected } R^2 = .56
\end{aligned}
$$

Faster rates of learning appear to take place in the beginning of the program, although the program length variable is not statistically significant. It is also noteworthy that the overall findings one would infer from equation (1B) are very similar to those one would infer from equation (1).

POOLING OF GRADE DATA

Stanford reading scores were available for grades 2, 3, 4, 5, and 6 in various combinations from project to project. The number of valid observations for single grade levels varied from 38 in grade 3 to 15 for grade 5. Grade 3 was the only grade for which more than 50 percent of the projects were represented. (A major reason for the large number of missing observations was that many projects changed test levels during the school year. This made their scores incomparable to the scores of projects that did not change levels.) Since achievement test scores are not necessarily comparable between grades (even when all scores are referenced to the norms by grade placement, as was done in this study), there is a possible objection to any procedure that pools data for different grades. On the other hand, if data were only used for the single usable grade, more than half of the performance data gathered in the study would have to be discarded. Discarding so much otherwise very useful information should be avoided if at all possible.

The solution that was adopted was to use pooled data if no apparent differences could be found among grade results after analyzing grade differences statistically. The test involved two steps. (1) First end score was regressed against beginning score for each grade to see if there were any discernible differences in this relationship by grade. There were not. (2) Then each grade was compared with grade 3 using a dummy variable for grade effect and covarying for beginning score. (It was not necessary to covary for program length since it was always virtually the same in the same school.) As an example of the procedure used, if there were 20 schools with scores for grades 3 and 4, the equation would have 40 cases and would be

$$\text{SCORE} = a_1 + a_2\ (\text{BEGIN SCORE}) + a_3\ ,$$

where a_3 is the coefficient of a dummy variable set equal to 1.0 if the observation were for grade 4 and zero otherwise.

The coefficients corresponding to a_3 for the four grade effects, with the t statistics for their standard errors, are:

	Coefficient	t
Grade 2	-0.08	0.42
Grade 4	-0.09	0.33
Grade 5	0.06	0.20
Grade 6	0.42	1.52

Since the coefficient for the grade 6 effect was large and almost significant statistically, grade 6 scores for 440 pupils for 19 projects were excluded. All the other grades were retained and a weighted pooled average of both end score and beginning score was constructed.

What are the possibilities that this procedure will lead to serious error? Differences in grade level effects could obtain because of different levels of resource inputs used at different grade levels or because of differences resulting from test construction. Since we have statistical evidence that there is no difference among the four grades used, the kind of error that could remain in the presence of this null finding would be offsetting errors; that is, increased resources might be used at a grade but be offset by the effect of test construction that biases gains downward. However, considerable care was taken in the interviews to check for differences in inputs by grade level, and there were not many instances in which they obviously differed (this is especially true with respect to grade 2, somewhat less true, perhaps, with respect to the findings for grades 4 and 5).

I doubt that this pooling procedure has led to serious error. Readers who disagree will have to use the findings presented in equation (4) and disregard the rest.

OTHER MINOR PROBLEMS IN CONSTRUCTING THE PERFORMANCE MEASURE

There were a number of relatively minor problems to overcome in using the Stanford Test Scores in this data set. First, it was found to be necessary to use the median performance scores as the measure of central tendency since in their reports some projects failed to include frequency distributions which would have been required to compute means. This allows for some bias, but careful investigation showed that the

difference between mean and median grade equivalents (many districts reported both) were non-existent or negligible.

A second problem arose because it was not possible to obtain summary scores for individual schools from some of the school districts. Twenty-two of the 42 school projects fell in this category. Half of the 22 had district reports where the school project being studied accounted for less than half of the pupils covered in the report. The method used to attempt to overcome this potentially serious data problem was to request the respondent to choose a school that was "closest to the district average" in performance. There was usually some such choice possible, and since district evaluation personnel often have a good feel for the performance levels of their project schools, the error introduced because of this mismatch was probably lessened considerably.

In equation (2B) the model is fitted to only those 31 projects where the mismatch problem was -- in terms of percentages, anyway -- relatively minor.

$$\begin{aligned}(2B)\quad \text{SCORE 25} = {} & -3.32 + \underset{(0.7)}{} + \underset{(1.9)}{4.35\ \text{PGMLENGTH}^{*}} - \underset{(1.7)}{.206\ \text{BEGIN 25}} - \underset{(0.2)}{.0040\ \text{PCTMIN}} \\ & + \underset{(3.1)}{1.48\ \text{SPECIEMS}^{*}} - \underset{(1.3)}{.022\ \text{PCTREGCR}} + \underset{(1.8)}{.089\ \text{TCHRPPIEMS}} \\ & + \underset{(0.8)}{.80\ \text{PLANHRS}} \\ & \text{SE Estimate} = 1.77 \\ & F(7,23) = 4.22 \\ & \text{Corrected } R^2 = .43\end{aligned}$$

Except for the less significant PLANHRS variable the equation is not greatly different from (1).

Finally, there was a problem with respect to the question of competing program outputs. The California Division of Compensatory Education requires that Title I projects teach both mathematics and reading. It was not possible to obtain comparable achievement data

on mathematics for 18 of the 42 projects,[1] however, and with this many missing observations it was simply not feasible to study mathematics programs directly. Instead a careful attempt was made to limit the study to resources going into reading.

WEIGHTING

A well-known problem to econometricians concerns the problem that regression equations fitted to sample populations where the expected error terms from properly specified models are not the same size along some important dimension of the analysis are not efficient. That is to say, other estimators can be found for which there is less error variance. There is one dimension in educational analysis such as that in this study where such expected error variance must surely differ, and that is program size. This is because mean scores of groups of pupils are used, and the expected error variance of means of small groups is greater than those for large groups, as everyone who has studied sampling theory knows.

An additional quirk to the analysis that has not been pointed out before in the educational input-output literature, however, is that there are two potential sources of randomness: a program effect applying to each student in the program, and a random effect that differs for each student coming because of the vagaries of achievement testing.[1] In symbols

$$u_{ij} = v_i + e_{ij} ,$$

where u_{ij} is the stochastic term for the jth student in the ith program, v_i is the effect of the ith program, and e_{ij} is a random term. The variance of the average test score across all students in the ith program depends on the number of students (size of program) because the sum of e_{ij} depends upon the number of students. The variance v_i

[1]Some districts did not include mathematics in their annual reports and others did not use the Stanford Mathematics tests.

[2]I owe this point to Joseph Newhouse.

due to program effects may or may not depend on size of program. (In point of fact, I would suspect that it does, since the law of large numbers works with teachers' effects and the like as well as with pupil performance on tests.) If v_i is independent of size of program, the question then becomes "How much of the total error term u_{ij} varies by program size and how much does not?" If a large percentage did not vary, it might be more correct not to weight, or to use only a partial weight.

It should be possible to get some insights about the propriety of weighting fully merely by performing the well-known test for heteroscedasticity. The projects were divided into four groups of 10, 11, 11, and 10 respectively ranged by sample size; the variance of the error term multiplied by a constant was computed for equation (1B). The result was as follows, where N = the number of pupils in the project whose scores were averaged:

1/N x 1000	Variance x 100
5.8	36.4
13.1	38.4
23.3	49.4
54.2	129.6

Variance obviously increases consistently with decreased sample size. If a regression line of variance is hand fitted to 1/N, the resulting line has a steep slope and an intercept fairly close to zero. This seems to indicate strongly that full weighting on the basis of sample size is proper.

REFERENCES

1. Alcaly, Roger, "Discussion," The American Economic Review: Papers and Proceedings, Vol. 61, No. 2, 1971, pp. 301-302.

2. Averch, Harvey, et al., "What Do We Know About Educational Effectiveness?" The Rand Corporation, Santa Monica, unpublished report.

3. Annual Report, Offices of Local Assistance, Department of General Services, State of California, Sacramento, 1970.

4. Bowles, Samuel, "Educational Production Function," in W. L. Hansen (ed.), Education, Income, and Human Capital, National Bureau of Economic Research, New York, 1970, pp. 11-61.

5. Bogue, Donald J., Principles of Demography, John Wiley & Sons, New York, 1969, pp. 770-771.

6. Carver, Ronald P., "Use of a Recently Developed Listening Comprehension Test to Investigate the Effect of Disadvantagement on Verbal Proficiency," American Educational Research Journal, March 1969, pp. 263-270.

7. Coleman, James S., and Nancy L. Karweit, Measures of School Performance, The Rand Corporation, R-488-RC, Santa Monica, July 1970.

8. Cronbach, Lee J., and Lita Furby, "How We Should Measure 'Change' -- Or Should We?" Psychological Bulletin, Vol. 74, 1970, pp. 68-80.

9. Far West Laboratory for Educational Research and Development, Individualized Instruction Information Unit, Berkeley, California, 1971.

10. Glass, Gene V., et al., Education of the Disadvantaged: An Evaluative Report on Title I, Elementary and Secondary Education Act of 1965, Fiscal Year 1969, University of Colorado, Boulder, 1970.

11. Gordon, E. W., and D. Wilkerson, Compensatory Education for the Disadvantaged, College Entrance Examination, New York, 1966.

12. Guszak, Frank J., "The Diagnostic Reading Teacher for the Disadvantage Child," in J. L. Frost and G. R. Hawkes (eds.), The Disadvantaged Child, Houghton Mifflin, Boston, 1970, pp. 361-378.

13. Halpin, Andrew W., and Dan B. Croft, "The Organizational Climate of Schools," Washington University, St. Louis, Mimeo, 1962, pp. 89.

14. Hanushek, Eric A., "The Education of Negroes and Whites," doctoral dissertation, M.I.T., 1968.

15. -----, The Value of Teachers in Teaching, RM-6362-CC/RC, The Rand Corporation, Santa Monica, December 1970.

16. Hawkridge, David, Albert B. Chalvpsky, and A. Oscar H. Roberts, A Study of Selected Exemplary Programs for the Education of Disadvantaged Children, Final Report, U.S. Department of Health, Education and Welfare, American Institutes for Research in the Behavioral Sciences, Palo Alto, 1968.

17. Hawkridge, David, Peggie L. Campeau, Kathryn M. DeWitt, and Penelope K. Trickett, A Study of Further Selected Exemplary Programs for the Education of Disadvantaged Children, U. S. Department of Health, Education and Welfare, American Institutes for Research in the Behavioral Sciences, Palo Alto, 1969.

18. Hawkridge, David, G. Kasten Tallmadge, and Judith K. Larsen, Foundations for Success in Educating Disadvantaged Children, Final Report, U.S. National Advisory Council on the Education of Disadvantaged Children, American Institutes for Research in the Behavioral Sciences, Palo Alto, 1968.

19. Jaeger, R. M., "The 1968 Survey on Compensatory Education," Paper presented at A.E.R.A. Meeting, 1969, Los Angeles.

20. Jensen, Arthur R., "How Much Can We Boost I.Q. and Scholastic Achievement?" Harvard Education Review, Vol. 39, 1969, pp. 1-123.

21. Johnston, J., Econometric Methods, McGraw-Hill, New York, 1960.

22. Katz, Daniel and Robert L. Kahn, The Social Psychology of Organizations, John Wiley & Sons, New York, 1966, Chapter 7.

23. Kiesling, Herbert J., California Compensatory Education Projects: A Draft Report on the First Part of an Economic Analysis of Compensatory Education Projects in California, The Rand Corporation, unpublished, Santa Monica, 1970.

24. -----, The Relationship of School Inputs to Public School Performance in New York State, The Rand Corporation, P-4211, Santa Monica, October 1969.

25. -----, Multivariate Analysis of Schools and Educational Policy, P-4595, The Rand Corporation, Santa Monica, March 1971.

26. McDill, Edward, Mary McDill, and Timothy Sprehe, Strategies for Success in Compensatory Education: An Appraisal of Evaluation Research, Johns Hopkins Press, Baltimore, 1969.

27. Posner, J., "Evaluation of 'Successful' Projects in Compensatory Education," U.S. Office of Education, Office of Planning and Evaluation, Occasional Paper No. 8, 1968.

28. Rapp, Marjorie, Evaluation as Feedback in the Program Development Cycle, P-4066, The Rand Corporation, Santa Monica. April 1969.

29. Ribich, Thomas I., "Comment on Paper by H. J. Kiesling," Conference on Education as an Industry, National Bureau of Educational Research, Chicago, June 4-5, 1971.

30. Statistical Report for Fiscal Year 1968: A Report on the Third Year of Title I, U.S. Office of Education, Department of Health, Education and Welfare, Washington, D.C., 1969.

31. U.S. Office of Education, Department of Health, Education and Welfare, Evaluative Report on Title I, Elementary and Secondary Education Act of 1965, Fiscal Year 1968, Washington, D.C., 1970.

32. Weisbrod, Burton A., "Comment on Paper by H. J. Kiesling," Conference on Education as an Industry, National Bureau of Educational Research, Chicago, June 4-5, 1971.