DOCUMENT RESUME

ED 090 323 UD 014 111

AUTHOR Gladkowski, Gerald

TITLE Evaluational Considerations of Compensatory Education Programs.

PUB DATE 73
NOTE 27p.

EDRS PRICE MF-$0.75 HC-$1.85 PLUS POSTAGE
DESCRIPTORS Compensatory Education; *Compensatory Education Programs; Control Groups; *Educational Planning; Educational Research; Evaluation Criteria; *Evaluation Techniques; Experimental Groups; Experiments; Formative Evaluation; Program Design; Program Development; *Program Evaluation; *Research Design; Research Methodology

ABSTRACT

The purpose of this paper is twofold: (1) to indicate some of the major weaknesses in the design and approaches to compensatory education programs, and (2) to recommend a more appropriate evaluational design. The second purpose deals specifically with a recommended evaluational procedure; i.e., the discussion centers around an account of what should be considered for inclusion if we are to adhere to the basic tenets of experimental research, and second, if we are to begin delineating relevant variables which affect the growth and development of impoverished children. On the basis of the discussion, the following factors are considered important in program planning: (1) the specific delimitation and delineation of a target area and sample within a specified geographic region. (2) After having decided upon the selection criteria, then a random sample would be selected from the population and assigned randomly to experimental and control groups. (3) The specific goals of each center schould be clearly delineated. (4) Evaluation procedures should be standardized and built into the program; that is, each center should employ similar measurement indices and schedules for gathering data. (5) Limit generalizations primarily to the specific geographical region. (6) Admit children in infancy, or a very young age. (7) Follow-up studies should definitely be included as part of the evaluation process. (8) Provide for "planned variations" between programs. (9) Provide sufficient time to "work out" many of the problems inherent in the program. (1) Utilize two staffs--one for research and one for every day implementation or treatment. (Author/JM)

EVALUATIONAL CONSIDERATIONS OF COMPENSATORY EDUCATION PROGRAMS

Gerald Gladkowski

The Pennsylvania State University

Evidence regarding the effectiveness of compensatory education is ambiguous, and is similar to the conclusions reached by others (for example, see Cohen, 1970; McDill, et al., 1969; Campbell, 1969). Much of the ambiguity revolves around two major areas, namely: (1) non-evaluational factors (e.g., size and scope of program, political interests), and (2) experimental or evaluational considerations (e.g., assignment of Ss to treatment groups). Obviously, neither of these factors are independent of the other. However, in this paper they will be treated as if they are in order to illustrate the many problems confronting such undertakings.

The purpose of this paper is twofold: (1) to indicate some of the major weaknesses in the design and approaches to compensatory education programs, and (2) to recommend a more appropriate evaluational design. The first purpose is included in order to provide a background of the major difficulties engendered by national assessments in general, and, more specifically, research designs which are primarily ex-post-facto, and which, by their very nature, create more problems regarding interpretation than they solve. The second purpose deals specifically with a recommended evaluational procedure; i.e., the discussion will center around an account of what should be considered for inclusion if we are to adhere to the basic tenets of experimental research, and second, if we are to begin delineating relevant variables which affect the growth and development of impoverished children.

BASIC PROBLEMS CONFRONTING COMPENSATORY PROGRAMS

Size and Scope: Cohen (1970) indicated that prior to 1964, educational evaluation had been primarily confined to small scale research in which the purpose of the study was generally limited to specific factors and typically involved a small budget and staff. However, after 1964, the federal government became involved in establishing broad educational programs which Cohen (1970, p. 213) perceived as differing from the previously conducted research in three important ways:

(1) They are social action programs, and as such are not focused narrowly on teachers' in-service training or on a science curriculum, but aim broadly at improving education for the disadvantaged.

(2) The new programs are directed not at a school or a school district, but at millions of children in thousands of schools in hundreds of school jurisdictions in all the states.

(3) They are not concerned and executed by a teacher, principal, a superintendent, or a researcher--they were created by a Congress and are administered by federal agencies from from the school districts which actually design and conduct the individual projects.

Without delineating all the questions and implications involved in the above, it is obvious that any large scale program will create many problems. For example, how does one effectively evaluate the specific effects upon approximately three million children spread out across the nation? Is it reasonable to evaluate on the basis of criteria related primarily to achievement programs directed at broad political, economic, and social changes? Should evaluation be decentralized despite the fact that national programs are involved? How does one determine the specific effects of any undertaking when the overall objectives for the program are determined nationally, but yet each local school district, or state, is responsible for implementation of the program? These are but a few of the questions that could be raised, and as Cohen (1970, p. 215) has stated, "In the social action programs, however, the political importance of information is raised to a high level by the broader political character of the programs themselves." The important point is that while the basic tenets of experimental research may be similar for evaluating both small and large scale programs (i.e., determining their effects) , the important difference lies in the character of the aims and organization of the program. Timpane (1970) and Campbell (1969, p. 410) reached similar conclusions with the latter stating that, "If the political and administrative system has committed itself in advance to the correctness and efficacy of its reforms, it cannot tolerate learning of failure. To be truly scientific we must be able to experiment. . ." For example, one would logically assume that some type of evaluational procedure would be involved in order to assess whether or not a program has been effective, but as Cohen (1970, p. 219) states:

> The mandate for evaluation--like many Congressional authorizations--lacked any enabling mechanism: responsibility for carrying out the evaluation was specifically delegated to the state and local education authorities who operated the programs. It was not hard to see, in 1965, that this was equivalent to abandoning much hope of useful program evaluation.

Campbell (1969) indicates that many feel we are at the point of continuing or discontinuing programs on the basis of assessed effectiveness, although he questions the validity of this attitude indicating that most ameliorative programs end up with no interpretable evaluation. Another example is the fact that Title I programs are funded on a formula grant type basis, in which the amount of money given to any educational district is based on how many poor children the district has enrolled in the schools, and not on how well the district may or may not educate. The actual implementation and evaluation of these programs are confounded by many non-evaluational considerations; for example, politically vested interests on various levels and the emotionally laden overtones of such programs. For a more detailed and complete discussion of other factors, one is directed to McDill, et al., (1969); Campbell, (1969); Cohen, (1970); and Timpane, (1970).

Variables: Another problem confronting compensatory education programs, specifically at the preschool level, has to do with the type of variables with which an investigator must cope. McDill, et al., (1969, p. 7) cites three important variables or factors which affect compensatory programs; namely, program effects or maturation, interactions of various socializing agencies, and technology.

Many programs are directed at preschool and elementary school children and are based, in part, on the belief that the earlier we begin assisting children of this age the more successful we may be. (For example, see Hunt, 1966) The problem this creates is that we have accumulated much more knowledge of the learning process and the effects of other variables upon children in the elementary school, relatively speaking, than those that affect preschool children. Only in recent years have efforts been made to study this much younger population. According to McDill, et al. (1969, p. 7), "Compensatory education or no compensatory education, we simply do not know much about how preschool children learn, and we know even less about disadvantaged learners." Because of this, it is difficult to determine whether the programs themselves are ineffective, or whether they are ineffective because of our inability to define the critical variables in order to assess the impact of the program. Campbell and Stanley (1966) discuss a related problem when they list maturation as a potential confounding variable which might possibly affect the internal validity of an experiment. They ask the question, "How does one distinguish between maturation and treatment effects in young children?" It should be indicated that compensatory education as a strategy is not in question, but, instead, the theoretical structure which supports the decisions that implement such a program. (Ginsburg, 1969, pp. 123-126) The present state of knowledge and the problems it creates for those interested in assessing the impact of various programs remains an obstacle to certainty in assessment. Generally speaking, researchers attempt to select one point in time as the input and another as the output, but research does not indicate if the two points are necessarily the most important in the life cycle of the individual, because it may be that the significant factors have occurred prior to the experimental treatment (a problem, by the way, in all research). It might be indicated that this is one reason why many recommend program implementation beginning in infancy, or at a much younger age than is presently included in such programs, hence increasing control over input variables. (See Boger and Ambron, 1969; Gladkowski, in press) The important point is that we do not actually know whether our programs have the effect they are designed to have, or as Zimilies (1969, p. 179) stated:

> The problem, then, is reduced to finding the appropriate inputs for achieving the desired output. While schematically this may appear to be an accurate analysis of the problem, it bypasses the critical intervening and mediating factor--the child. Nowhere does one find a description of the four-year-old child, a developmental analysis of the personality and cognitive functioning of children at this age level, or a statement of their primary areas of conflict, typical modes of resolution, and principle spheres of development.

Interaction between socializing agencies represents another important source of difficulty for evaluation. This problem revolves around the fact that education (in the broadest sense) does not take place exclusively in the schools. A child may be involved in a formal educational program for six hours per day, but what about the other eighteen hours? Does the remainder of the time outside the program cancel any potentially positive effects that might have occurred during the treatment? Is there an optimum amount of time spent in school which could be effective? What effect do significant others have upon the child, e.g., peers and parents? The answers to these questions are, of course, not available at the present time, although they are questions which will eventually need answers if we are to identify and assess the effectiveness of our programs. More will be reported regarding this uncontrolled source of variance later in the paper.

Gordon (1970) presented an excellent overview of various attempts to assist disadvantaged segments of our society in which he provides a brief synopsis concerning the areas of concern and directions for approaching the problems in program implementation for the disadvantaged. Much of the difficulty of explanation and interpretation of the various positions arise due to the confounding of factors in an attempt at delineation. For example, it has been shown that as Southern Blacks move North, their achievement levels increase. The question arises, however, as to whether this is due to the impact of the school, selective migration, non-school environmental conditions, the interaction of these factors, or others not yet investigated. The interaction of many factors increases the complexity of attempts at explaining any outcome of an intervention effort. (For example, see Grotberg, 1969)

According to McDill, et al., (1969), if one had a firm idea of the relevant variables important to any program design, one would still be faced with the question of measurement. How much can we rely on our measurement devices to give us the data we need for evaluating outcomes? The difficulty arises at all levels, but even more so at the preschool level because of the relative lack of measurement data concerning this age range with it generally acknowledged that, the younger the child, the more inaccurate our measurement devices are likely to be. For example, if a child were tested at age two on one of the standardized infant scales available, we would not expect as nigh a correlation with later achievement as we would if we were to administer the test at age seven and correlate it again, at say, age ten. McDill, et al. (1969) indicates that while the state of development regarding cognitive dimensions is still "primitive", the picture is even more depressing when one considers the affective domain. (See Wick and Beggs, 1971; Cronbach, 1960; Mehrens and Lehmann, 1970).

Specific Factors: The discussion presented above concerned itself primarily with general factors affecting evaluational research, whereas this section will delineate some of the more specific research problems relating to compensatory education programs. In addition, alternatives to the specific weaknesses cited will be presented, with the paper concluding with a listing of the factors that should be considered in a well-designed experimental effort.

One of the primary difficulties inherent in compensatory programs has been an obvious lack of control over relevant variables ranging from non-comparable groups for comparison, (no control groups in certain instances), to the interaction effects of the environment. (McDill, et al., 1969) For example, the evaluation of Project Head Start contained many factors which were uncontrolled in the design. First, randomly selected experimental or control groups were not used but instead an ex-post-facto-design in which the controls were selected and matched after the experimentals had already received the treatment constituted the basis for the evaluation. This, of course, makes it impossible to determine the specific effects of the program and thus violates one of the basic tenets of experimental research. It should be indicated that the evaluators of Project Head Start did randomly select the centers for the study, but this was invalidated by many previously cited weaknesses inherent in the assessment of various local programs, with the following factors being cited as representative of these weaknesses: (Westinghouse-Ohio, 1969)

1. Lack of comparability among separate and independent studies because of different enrollment criteria, program treatments, design, instrumentation, and schedules for gathering data.
2. In some cases the absence of any comparison group.

3. Too few cases, frequently only those enrolled at a particular center.

4. Geographical restrictions to local or regional groups.

On the basis of these difficulties, selecting a "random" sample of an already biased or non-comparable sample does not eliminate the sources of bias. (See Harvard Educational Review, 1970).

Second, there were no uniform or standardized procedures adhered to between various programs to insure that the evaluation would be attempting to assess those factors which programs shared in common. For example, the various centers employed somewhat different goals, treatments, and program procedures, thus masking between and within center differences. Some centers were in operation for two hours per day whereas others were in operation for four hours; some centers were only in operation for two months whereas others were in operation for eight or nine months out of the year. (See Cohen, 1970 and McDill, et al., 1969) Despite these differences the programs were all evaluated as if they were similar; however, there is no way of ascertaining which specific centers were relatively "successful" as compared to those which were not. Regarding this masking effect, Cohen (1970, p. 226) stated, "The problem, then, is not only to identify what the programs deliver, but also to systematically experiment with strategies for affecting school outcomes....The movement toward experimentation presumes that the most efficient way to proceed is systematic trial and discard, discovering and repeating effective strategies." Others who hold similar views regarding "planned variations" include Smith and Light (1970) and Campbell (1969). This approach was not employed in the Head Start Project although the evaluative team did recommend this for future consideration.

In the assessment of Project Head Start, the emphasis was on "overall" effectiveness of the program, disregarding those centers which might have been particularly effective. What this would mean in practice is that if a center (or certain aspects of a center) were found to be particularly effective, then one could further investigate it in order to determine how it differs from the other centers or programs in its operation. If significant differences were detected, then other centers could be organized in which the best features of proven programs could be incorporated, as well as the fact that presently operating programs could thus be modified.

Other weaknesses which contributed to the overall evaluational efforts included lack of uniformity across the various centers regarding such matters as the use of the same indices of measurement, objectives of the program, and the selection criteria of Ss for treatment and control groups. This uniformity had not been accomplished in many of the programs, because, in part, the local programs were permitted the freedom to not only evaluate their own programs but also to decide upon a specific implementation course. As stated by Cohen (1970, p. 227), "The Office of Education. . . . does not require that the same tests be used in all Title I projects; indeed, it does not require that any tests be used." In order for an appropriate evaluation to be undertaken, such matters as this must be considered before the implementation of the program; thus obviating later problems arising regarding interpretation of the results.

Many of the weaknesses inherent in the experimental designs are those related to internal validity; that is, those factors associated with the question: Did the experimental treatments make a difference in this specific experimental instance? (See Campbell and Stanley, 1966) With so many weaknesses in evidence, it is virtually impossible to answer this question. Hence, the studies undertaken to date are of very limited scientific value in determining whether or not the programs were effective. The following comprises the major weaknesses of compensatory evaluations and would thus form a rather formidable list of competing alternative hypotheses to any research undertaking:

1. Lack of comparable groups, and, in some cases, no control groups at all.

2. No planned variation in programs in order to assess both

2. (Continued) within and between center differences.
3. Lack of random selection and/or assignment of Ss to treatment and control groups.
4. Lack of clear-cut criteria for inclusion into the program.
5. Lack of clearly specified objectives.
6. Non-comparable data, i.e., different indices of measurement.

In lieu of the above, one needs to ask: What factors should be included for a more rigorous evaluational procedure? The position this paper will advance is based primarily upon the recommendations of Campbell (1969), Campbell and Erlebacher (1970), and McDill, et al. (1969) in which they recommend that future intervention programs adhere to the basic tenets of experimental research and closely approximate a "true" experimental design. As stated by Campbell (1969, p. 410), "We must be able to advocate without that excess of commitment that blinds up to reality testing." If we are interested in delineating the specific effects of variables upon subsequent development in compensatory education programs, then we should attempt to cope with the problem by employing the most accepted and theoretically sound procedures possible (however imperfect they may be).

CONTROL FACTORS

Experimental and Ex-Post-Facto Studies: One of the most important differences between experimental and ex-post-facto research is control. In the former, the logic of controlled experimentation produces data which predicts Y as a function of X; whereas in the latter, we begin with Y and then retrospectively seek to define X. While ex-post-facto studies have value, the investigator is placed in the unenviable position of asserting without the certainty of cause and effect, because the X has already occurred, with Kerlinger (1967, p. 371) citing the following weaknesses of such studies as:

1. The inability to manipulate independent variables,
2. The lack of power to randomize, and
3. The risk of improper interpretation.

Many of the compensatory programs undertaken to date would be classified as ex-post-facto and no doubt contribute to the ambiguity of the results reported. Certainty, of course, is never reached; it is only approximated even in experimental research, although it is generally recognized that one can place considerably more reliance in the findings of adequately controlled experimental investigations. (see Hays, 1963 or Edwards, 1968)

Given this distinction between experimental and ex-post-facto research what factors should be included in an evaluational design in order to approximate more closely an experimental approach? The following principle to be described below provides an excellent account of the purposes of research design and statistical analyses while also suggesting factors which should be considered in the planning of any evaluation. After having presented this account, a discussion of some of the more important variants or derivatives of the principle will be discussed.

Maximinicon Principle: According to Kerlinger (1967, p. 280) the main technical function of research is to "control variance," so in essence, "a research design is, in a manner of speaking, a set of instructions to the investigators to gather and analyze his data in certain ways and is therefore a control mechanism." The statistical principle behind this mechanism is what is referred to by Kerlinger as the "maxminicon" principle; that is, the maximization of experimental variance, the minimization of error variance, and the control of extraneous systematic variance. Before stating certain procedures for utilizing this principle, it would be advisable to clarify the sources of variance. In an experiment it is the dependent variable measures that are analyzed. From this analysis we can infer that the variances present in the total variance of the dependent variables are due to the manipulation and control of the independent variables. (Kerlinger, 1967, p. 282)

Maximization of Experimental Variance: In most research, one of the investigators major concerns is to maximize the experimental variance. This variance can be either "assigned" or "active", depending upon the control the investigator has over the variable. For example, sex is an assigned independent variable, because it is constant within the same person; whereas, methods of instruction would be an active independent variable, because the investigator can control or manipulate the actual instructional method employed. In order to maximize the variance, it would be advisable to pull the methods (treatments) apart as much as possible, i.e., make them as different as possible, and in this manner the experimenter is permitting the variance of a relationship to show itself apart from the total variance.

Control of Extraneous Variance: The control of extraneous variables refers to the influence of independent variables extraneous to the purposes of the study being minimized, nullified, or isolated. According to Kerlinger (1967, p. 284) the variance of such variables is in effect reduced to zero or near zero. That is, it is separated from the variance of other independent variables of concern. There are primarily four ways in which one can control extraneous variance; namely, elimination of the variable as a variable, randomization, build control into the design as an independent variable, or matching. Of the four, the one most often recommended is randomization. (See almost any text on experimental design and research, e.g., Campbell and Stanley, 1966; Hays, 1963; Kerlinger, 1967; or Edwards, 1968, for a more complete discussion.

Theoretically, randomization is the only method of controlling all possible extraneous variables with this concept being one of the most commonly accepted dictums of experimental research. In practice, however, adequate randomization has seldom been achieved. Campbell (1969) and Campbell and Erlebacher (1970) reiterate the importance of future social reform programs employing the random selection and assignment of Ss to control and experimental groups. This principle, if adhered to, does not mean that the groups are equal in every conceivable way, but that the probability of their being equal is much greater than the probability of their not being equal. For example, the environment is an important source of interference in any study, and, in the past, has probably contributed much to the confounding that has occurred in various programs, but yet, is uncontrolled in most compensatory programs. The principle resulting from this concept was posited by Kerlinger (1967, p. 285) as: "Whenever possible to do so, randomly assign conditions and other factors to experimental and control groups." Although this principle engenders certain ethical considerations, the present writer adopts a rather simplistic rationale; namely, if X dollars are available and Y persons need assistance, then you help those you can. In other words, X is generally consistently less than what is needed so the persons who need assistance will not all be included in the program anyway. If this is the case, then why not randomly offer assistance. This would appear preferable to having the political considerations enter into the process.

Minimization of Error Variance: The third aspect of the principle described by Kerlinger is the minimization of error variance; namely, the variability of measures generated by random fluctuations which have a tendency to balance each other so that their mean is zero. This is contrasted with systematic variance, or the tendency for measures to vary consistently in one direction or another. The determinants of error variance include those due to individual differences and measurement. The minimization of error variance includes two principle aspects: the reduction of errors of measurement through controlled conditions and an increase in the reliability of the measures. The more uncontrollable the conditions of the experiment, the more the determinants of error variance can operate.

In a well designed experiment, the various factors which may influence the outcome of the experiment, and which are not themselves of concern, must be controlled if valid conclusions are to be drawn concerning the results of the experiment. Edwards (1968) discussed these factors emphasizing that these conclusions are derived from the structure of the experiment and the nature of the controls exercised. They do not come from the test of the null hypothesis. The statistical test employed indicates only the probability of a particular result based upon the statistical hypothesis tested, namely, that chance alone is determining the outcome. If the experimenter rejects the null hypothesis, he must still examine the structure of the experiment and the nature of his experimental controls in making whatever explanations he does make concerning why he obtained the particular result. With this clarification, it becomes extremely important to consider other factors which might influence the particular results, and which if not considered could possibly serve as competing alternative hypotheses to the results obtained.

Sample Delimitation and Generalizability: One such factor of importance is the specific delimitation of the sample to be employed in the study, that is: What type of individual(s) will one consider for inclusion into the program? This question was confused in some of the previously conducted compensatory education programs as indicated by the fact that the criteria for admission into the programs varied by geographical region, as well as between centers within regions, and hence confounded an adequate comparison between centers. Equally as important is the specification of the control group so that, again, adequate comparisons can be made. In other words, any program would call for the specific delimitation of a target area and population within predesignated regions. For example, all families residing within the city of Evanston, Illinois, who earn below X number of dollars, and possess no more than Y number of children are eligible for admission. As stated previously, this would be done on a random basis so that each subject within the specified area had an equal opportunity for selection into the treatment and control groups. If this is done between centers, assuming there are more than one, then we can be more certain of comparability and hence should reduce one competing alternative hypothesis: namely, biases resulting from differential selection of respondents for the comparison groups.

Limiting generalizations primarily to the specific geographical region and/or sample also insures potential generalizability within an area, although it would be possible to generalize beyond the specific area. (See Edgington, 1969, for a more complete discussion of extrapolations beyond the actual sample employed). By delimiting extrapolations to more manageable geographic regions, one could be more reasonably assured of applicability. Campbell and Stanley (1967, p. 17) offered a caveat regarding external validity when they stated that, "*Logically*, we cannot generalize beyond these limits, i.e., we cannot generalize at all. But we do attempt generalization by guessing at laws and checking out some of these generalizations in other equally specific but different conditions." One of the implications of this caveat is that of replications over time.

Standardization of Indices: Evaluational procedures should be built into the program prior to implementation as well as the standardization of the measurement indices. By standardizing procedures, it will be much easier to administer various measurement devices to be used in the evaluational procedure as well as to designate the specific times this is to be accomplished. For example, one might administer two indices every year to both experimental and control groups at approximately the same time, which could be specified before the program is undertaken. The schedules for collecting data would thus be uniform both within and between programs for both experimental and control groups. This procedure would also reduce competing alternative hypotheses of the results obtained, such as the non-comparability of data, and would thus increase the control dimension. If programs employ similar goals, treatments, and measurement indices, then the masking between and within programs should be considerably reduced. (See Smith and Light, 1970)

Another related suggestion would be to include an evaluation team from outside the specific geographic region to conduct subsequent assessments. It would also be advisable to have one group of observers for both experimentals and controls and preferably where the observers do not know to which group the child belongs. In this manner both groups could be randomly assigned to testing sessions in which the test could be individually administered at approximately the same time of the day. With young children, someone close to the child may be needed for assistance, but this should have no effect if the testing team does not know the staff of the center, et cetera. (See Campbell and Stanley, 1967; Kessen, 1969; and Wick and Beggs, 1971, for a complete discussion of the various evaluational considerations). The point being that we could improve this dimension by planning a strategy before implementation.

Multivariate-Experimental Longitudinal Approach: Shulman (1970) recommended varied research strategies for those interested in investigating the effects of the educational process, one of which was a multivariate-experimental longitudinal approach. This approach is similar to the one recommended in this paper, although Shulman was specifically concerned with the educational process in the classroom per se, whereas many programs in compensatory education represent a broader concern of which the classroom is but one sub-part. Despite this basic difference, many of the underlying principles remain similar. Shulman (1970, p. 387) described what he believed to be the common characteristics within a classroom situation as follows:

1. they involve the attempts to modify or manipulate a setting. . . to bring about desired changed in a learner;
2. they take place over relatively extended periods of time;
3. they involve the simultaneous input of multiple influences and the likely output of multiple consequences--some predicated, others not, and;

4. they are characterized by variability of reaction to ostensibly common stimuli, that is, not all learners learn equally or react similarly to specific acts of teaching.

Shulman (1970, p. 388) further lists four factors which would characterize "ideal" research, particularly if it is to be consistent with the four situational factors listed above, namely:

1. experimental
2. longitudinal
3. multivariate at the level of both independent and dependent variables, and consistent with that,
4. differential, in that intereactions of the experimental programs with the students' entering individual differences are treated not as error variance, but as data of major interest in the research.

Another recommendation is that programs be designed so that different experimental groups enter the treatment phase at different stages, and, in essence, is another way of implementing a "planned variations" approach. One way of accomplishing this, for example, would be to admit children of varying ages into a program in order to determine the effects upon Ss at different ages in order to answer such questions as: Is there an optimal age at which intervention should be begun? Is there an interaction effect between children of varying ages in the program? Does the program work better with the most "disadvantaged" segment of the population? Does it work better with families with a certain number of children? The various combinations are virtually unlimited and would contribute tremendously to our knowledge regarding specific effects upon subsequent behavior.

Planned Variations: It was noted previously that despite the many differences between compensatory education programs (e.g., admission criteria, length of time in operation, and different treatments), the programs were evaluated "as if" they were similar; however, there was no way of ascertaining which specific centers were relatively successful as compared to those which were not. Regarding this masking effect, Cohen (1970, p. 226) stated that, "The problem, then, is not only to identify what the programs deliver, but also to systematically experiment with strategies for affecting school outcomes...the movement toward experimentation presumes that the most efficient way to proceed is systematic trial and discard, discovering and repeating effective strategies." Such an approach was not employed in the Head Start Project. The evaluation emphasized the "overall" effectiveness of the program, disregarding centers which might have been particularly effective. In practice, if a center (or certain aspects of a center) was found to be particularly effective than one should further investigate it in order to determine how it differs from the other programs in operation. If significant differences were identified, then other centers could be established in which the best features of proven programs could be incorporated. In addition, presently operating programs could thus be modified in which the evaluation would concern itself with both within and between center differences.

Smith and Light (1970, p. 9) noted that a program may be partially successful in certain areas and not so in others, but this becomes of little or no concern if one can go back and support those weaker areas. One must recognize, however, the tremendous difficulties of trying to maximize simultaneously goals in more than one group. Other factors cited by the authors included consideration of the impact on the individual child rather than a dependence on an average for the entire group which might mask any specific effect for an individual. (See Wick and Beggs, 1971, Chapter Three) In addition, the importance of the program being replicated, monitoring for unintended results, the concern for not only successful but unsuccessful outcomes in order to assist in future plannings of programs, and the fact that control is important are other factors to be considered in evaluating program effects. Smith and Light (1970, p. 11) further recommend that we also concern ourselves with within-center differences, because they believe the relevant question to be: Which of the program centers worked well for reasons which are known and which can be reestablished in any future program centers?

Interaction: An emphasis, or more accurately a re-emphasis, has begun to be directed at the situational contexts that confront an individual and its effects upon behavior. This concern has been termed "inter-action analysis" and is typified by the work of such individuals as Amidon and Hough (1967) and Flanders (1964). Mitchell (1969, p. 697) claims that in spite of such evidence, the current situation is much the same as it was in 1955, when Rotter stated the following:

> In the half century or more that psychologists have been interested in predicting the behavior of human beings in complex social situations, they have persistently avoided the controvertible importance of the specific situation on behavior. . . So they have gone from faculties and instincts and sentiments to traits, drives, needs, and their inter-action of these within the individual, producing schema of personality organization and classification of internal states but ignoring an analysis of psychological situations in which human beings behave.

As Mitchell (1969, p. 698) states, "if the person-environment interaction is critical for understanding and predicting human behavior, it is equally apparent that this interaction can only be defined effectively in multivariate terms." This position is similar to that cited by Shulman (1970), with the important question being the research methodology appropriate to give meaning to such conceptions. The important point is that interaction has been a source of difficulty in compensatory programs. There are many facets to the study of interaction found in the studies of those who prefer the laboratory to those who prefer field or natural environments.

Cronbach (1957) has presented a paradigm for the delineation of specific aspects of interactions between an individual's aptitudes with a particular class of environmental variables such as instructional methods or treatments. Cronbach (1957, p. 680) discussed his position on interaction as follows:

> Applied psychologists should deal with treatments and persons simultaneously. Treatments are characterized by many dimensions: so are persons. The two sets of dimensions together determine a payoff surface. For any practical problem, there is some best group of treatments to use and some best allocation of persons to treatments. We can expect some attributes of persons to have strong interactions with treatment variables. These attributes have far greater practical importance than the attributes which have little or no interaction.

Cronbach recommended varied approaches based upon individual differences of the learners and is thus similar to the approach recommended by others. (For example, see Bloom, 1968; Mathis, et al., 1970; Bracht, 1970). The concern is not for the "best approach" but rather for varying approaches based upon given characteristics of the learner. While there are many strategies one could employ, the important point is that the use of varied strategies would lend itself to such analyses, because it would be relatively easy to vary instructional and program alternatives. While programs are intended for a certain specified segment of the population, there is no reason why successful approaches (if detected) could not be utilized for other populations. It should be noted that interaction effects are difficult to interpret, though not necessarily to detect, although this does not speak against our attempts at their assessment and interpretation. The implications are that the interactions between individuals and their environments are important, and always have been, but the present state of knowledge regarding such phenomena is just beginning to be developed. Mitchell (1969, p. 704) provided perhaps the best advice when he stated that, "conceptualizations in multivariate terms is not likely until the results of simpler investigations are in evidence begins to accumulate that the appraoch is fruitful."

Action and Research: An often overlooked problem which can affect research of the type recommended is the essential differences between action and research. McDill, et al, (1969) noted that the emphasis today is placed on not only immediate but successful modes of social action, with Hawkridge, et al. (1968, p. 15) stating that:

> Action and research are to some extent incompatible. The first seeks to guarantee a predetermined outcome; axiomatically it spares no effort and is entirely dependent upon the existing store of knowledge and information; time is of the essence. Research, on the other hand, is often slow; unless it deliberately and selectively restricts the scope of action, it may seriously handicap the attempt to add new knowledge to the existing store.

Kessen (1969) and Campbell (1969) voice similar sentiments, taking the position that we should investigate the problem much as we would any research problem. By removing an important source of variance, more accurately converting what was once error variance into systematic variance, one can begin evaluation at periodic intervals realizing the need to "work out" many problems inherent in any undertaking of the nature proposed in this thesis. Campbell and Stanley

(1967, p. 3) cogently described the "spirit" of research undertakings when they stated:

> The experiments we do today, if successful, will need replication and cross validation, at other times and under other conditions before they can become an established part of science, before they can be theoretically interpreted with confidence. . . Thus we might expect. . . . an experimental outcome with mixed results, or with the balance of truth varying subtly from experiment to experiment. The more mature focus--and one which experimental psychology has in large part achieved--avoids crucial experiments and instead studies dimensional relationships and interactions along many degrees of the experimental variables.

<ins>SUMMARY</ins>: The previous discussion has attempted to include those factors which should be controlled in order to reduce their subsequent effects as competing or rival alternative hypotheses. That is, if one has randomly selected and assigned Ss to treatment and control groups and clearly delineated a sample, then a competing alternative hypothesis of non-comparability of samples is considerably reduced. Many of the research endeavors undertaken are dependent upon this premise of minimizing extraneous sources which might possibly influence the results of an experiment, with the important factor being that of control.

Recent attempts in the field of compensatory education have been beset by a myriad of factors interacting simultaneously, thereby confounding both process and expected results. (For example, see Gordon, 1970) Another way of viewing this process is that we have been able to assess the "output" variables but have had extreme difficulty in specifically assessing and delimiting the input dimensions. (For example, see Grotberg, 1969)

On the basis of the preceding, the following factors should be considered in program planning:

1. The specific delimitation and delineation of a target area and sample within a specified geographic region. For example, a locale might decide that all the families who fall below a designated income level and who possess X number of children are available for inclusion into the program.
2. After having decided upon the selection criteria, then a random sample would be selected from the population and assigned randomly to experimental and control groups. Those who do not want to participate will, of course, be permitted not to do so, but a record should be maintained on these individuals as well as those who begin the program and subsequently drop out (experimental mortalities). An assumption underlying this is that the program will be explained to the prospective population--both the program and rationale for employing a random sample.
3. The specific goals of each center should be clearly delineated, preferably in behavioral objective form when possible. This should be done for the program as a whole as well as the individual subparts.
4. Evaluation procedures should be standardized and built into the program; that is, each center should employ similar measurement indices and schedules for gathering data. Multiple independent and dependent measures should be employed and administered at approximately the same time to both treatment and control groups.
5. Limit generalizations primarily to the specific geographical region. It would be possible to generalize beyond the specific area although, as always, with extreme caution. Be delimiting extrapolations to a specified region (and sample), one could be reasonably more assured of applicability and is analagous to a

"small steps" approach employed by many experimental psychologists.

6. Admit children in infancy, or a very young age, thereby reducing the input-output dimension as a competing alternative hypothesis. Of course, this is a recommendation not a prescription, because it might be more advantageous to vary the ages in order to determine, for example, the optimal age for admission as could be done with the criteria for determining which families are elibible.
7. Follow-up studies should definitely be included as part of the evaluation process.
8. Provide for "planned variations" between programs. For example, you might have two centers which are exact replicas of each other, and two others which are also replicas although different from the first set. In this way, one could compare the "overall" effectiveness, between the four centers as well as the between and within center differences. This would hopefully provide adequate comparisons which could then be used to identify the most successful as well as the lease successful features of various programs.
9. Provide sufficient time to "work out" many of the problems inherent in the program, i.e., emphasis on formative evaluation.
10. Uitlize two staffs--one for research and one for every day implementation or treatment.

REFERENCES

Amidon, E.J., and Hough, J.B. Interaction Analysis: Theory, Research and Application. London: Addison-Wesley Publishing Company, 1967

Bloom, B.S. Stability and Change in Human Characteristics. New York: John Wiley and Sons, Inc., 1964.

Bloom, B.S. "Learning for Mastery." Evaluation Comment, 1 (2). Center for the Study of Evaluation of Instructional Programs, U.C.L.A., May, 1968.

Boger, R.P., and Ambron, S.R. "Subpopulational profiling of psychoeducational dimensions of disadvantaged preschool children," in Grotberg, E. (ed.), Research Related to Disadvantaged Children. Princeton, N.J.: Educational Testing Service, 1969.

Bracht, G.H. "Experimental factors related to aptitude-treatment interactions." Review of Educational Research, 1970, 40(5), 627-645.

Campbell, D.T. "Reform as experiments." American Psychologist, 1969, 24(4), 409-429

Campbell, D.T., and Erlebacher, A. "How regression artifiacts in quasi-experimental evaluations can mistakenly make compensatory education look harmful," in Hellmuth, J. (ed.), Compensatory Education: A National Debate, Volume III of the Disadvantaged Child. New York: Bruner/Mazel, 1970.

Campbell, D.T., and Stanley, J.C. Experimental and Quasi-Experimental Designs for Research. Second Printing. Chicago: Rand McNally and Company, 1966.

Cohen, D.K. "Politics and research: Evaluation of social action programs in education." Review of Educational Research, 1970, 40(2), 213-258.

Cronbach, L.J. Essentials for Psychological Testing. New York: Harper and Row, 1960.

Cronbach, L.J. "The two disciplines of scientific psychology." American Psychologist, 1957, 12, 671-684.

Edgington, E.S. Statistical Inference: The Distribution Free Approach. New York: McGraw Hill Book Company, 1969.

Edwards, A.L. Experimental Design in Psychological Research. Third Edition. New York: Holt, Rinehart and Winston, 1968.

REFERENCES (Continued)

Flanders, N.A. "Some relationships between teacher influence, pupil attitudes, and achievement," in Biddle and Ellena (eds.), Contemporary Research on Teacher Effectiveness. New York: Holt, Rinehard, and Winston, 1964.

Ginsburg, H. "Review section," in American Educational Research Journal, 1969, 6(1), 123-126.

Gladkowski, G.J. "Another look at compensatory education," in Research in Education. (in press).

Gordon, E.W. "Introduction," Review of Educational Research, 1970, 40(1), 1-12.

Grotberg, E. Critical Issues in Research Related to Disadvantaged Children. Princeton, N.J.; Educational Testing Service, 1969.

Harvard Educational Review. "Equal educational opportunity." Winter, 1968, 38(1).

Hawkridge, D.G., Chalupsky, A.B., and Roberts, A. A Study of Selected Exemplary Programs for the Education of Disadvantaged Children. U.S. Office of Education Final Report, Project 089013, 1968.

Hays, W.L. Statistics. New York: Holt, Rinehart and Winston, 1963.

Hunt, J. McV. "The Psychological Basis for Using Preschool Enrichment as an Antidote for Cultural Deprivation," in Hechinger, F.M. (ed.), Preschool Education Today. New York: Doubleday and Company, Inc., 1966.

Kerlinger, F.N. Foundations of Behavioral Research. New York: Holt, Rinehart and Winston, Inc., 1967.

Kessen, W. "Early learning and compensatory education: Contribution to basic research," Educational Resources Information Center Document Resume, ED 036 318, August, 1969.

Mathis, B.C., Cotton, J., and Sechrest, L. Psychological Foundations of Education: Learning and Teaching. New York: Academic Press, 1970.

McKill, E.L., McDill, M.S., and Sprehe, J.T. Strategies for Success in Compensatory Education: An Appraisal of Evaluation Research. Baltimore, Md.: The Johns Hopkins Press, 1969.

REFERENCES (Continued)

Mehrens, W.A., and Lehmann, J.J. Standardized Tests in Education. New York: Holt, Rinehart and Winston, Inc., 1969.

Mitchell, J.V. "Educational challenge to psychology: The predication of behavior from person-environment interactions." Review of Educational Research, 1969, 39(3), 695 721.

Shulman, L.S. "Reconstruction of educational research." Review of Educational Research, 1970, 40(3), 371-396.

Smith, P.V., and Light, R.J. "Choosing a future: Strategies for designing and evaluating new programs," Harvard Educational Review, 1970, 40(1), 1-28.

Timpane, P.M. "Educational experimentation in national social policy," Harvard Educational Review, 1970, 40(4), 547-566.

Westinghouse Learning Corporation-Ohio University. "The impact of headstart: An evaluation of the effects of headstart on children's cognitive and affective development. ERIC Document Resume, Ed 036 321, June, 1969. (Executive Summary).

Wick, J., and Beggs, D.L. Evaluation for Decision Making in the Schools. Boston: Houghton Mifflin Company, 1971.

Zimilies, H. "Review of Martin Deutsch's book The Disadvantaged Child." Harvard Educational Review, 1969, 39, 177-180.