DOCUMENT RESUME

ED 118 603 TM 005 089

AUTHOR Johnson, Thomas J.; And Others
TITLE Causal Modeling in Educational and Social Program Evaluation.
PUB DATE [Apr 75]
NOTE 18p.; Paper presented at the Annual Meeting of the American Educational Research Association (Washington, D.C., March 30-April 3, 1975)

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage
DESCRIPTORS *Evaluation Methods; *Models; Preschool Programs; *Program Evaluation; Research Methodology
IDENTIFIERS *Causal Modeling

ABSTRACT

Educational and social programs often develop from a weak or imprecise conceptualization relating the program's system of input variables to its claimed outcomes. Evaluation personnel can contribute both to the final development of a program and to the fair evaluation of such programs by learning to formally characterize programs and to construct causal models of them. The evaluation effort represents an attempt to determine the correctness of the program's existing conceptualization, and if properly carried out, permits the developer/sponsor to strengthen, add, or delete components which are found to be nonfunctional. In this paper, the authors discuss the concept of causal model building and illustrate their ideas with an example of how causal model construciton procedures were used to assist in the evaluation of a complex early childhood program. (Author)

CAUSAL MODELING IN EDUCATIONAL AND SOCIAL PROGRAM EVALUATION

Thomas J. Johnson, Boston University
David R. Buckholdt, Marquette University
Daniel E. Ferritor, University of Arkansas

Educational and social programs often develop from a weak or imprecise conceptualization relating the program's system of input variables to its claimed outcomes. Evaluation personnel can contribute both to the final development of a program and to the fair evaluation of such programs by learning to formally characterize programs and to construct causal models of them. The evaluation effort represents an attempt to determine the "correctness" of the program's existing conceptualization, and if properly carried out, permits the developer/sponsor to strengthen, add, or delete components which are found to be non-functional. In this paper, the authors discuss the concept of causal model building and illustrate their ideas with an example of how causal model construction procedures were used to assist in the evaluation of a complex early childhood program.

Paper presented at meetings of American Educational Research Association, Washington, D. C., April 1-4, 1975.

Objectives

The purpose of this paper is to discuss the concept of causal model building as applied to educational and social program evaluation. This objective will be pursued through a presentation of the concept of causal modeling as reflected from formal program characterization procedures, a review of current prominent conceptions about educational and social program evaluation and their relation to causal modeling, the presentation of several examples of the use of causal model building in evaluation studies, and an examination of the policy implications for the evaluator.

Introduction

In recent years, the demand for evaluation activities has grown as the public, governmental agencies, and social scientists themselves have begun to require more accountability from tax-supported social programs. Thus, initially (and naively) evaluation was undertaken to provide a simple overall assessment of a program's ability to achieve its specified objectives or outcomes. This conception of evaluation proved to be limited, however, as the experience of evaluators revealed that the task was not so simple, the evidence often hard to come by, and the conclusions not so clear cut. New dimensions began to be added to the scope of an "evaluation," including concern about the merits of the objectives, sensitivity to unanticipated effects of programs, the information needs of different audiences, the different evidence requirements of formative and summative evaluations, and the analysis of political contexts in which evaluations are done. The adequacy of the conceptualization of the "program," the thing to be evaluated, however, has not been treated as an equally important dimension in the evaluation effort.

The conceptualization of a program involves the construction of a model of the important operational components of the program and the specification of the expected linkages between program components and between program components and program outcomes. From this point of view, an ideally conceptualized program is one which specifies how the system of input variables called "program" operates to effect the outcome variables claimed for it. This is the causal model. The <u>program</u> itself is then conceived of as merely a vehicle for delivering the system of input variables. The causal model, embedded in the program, is a falsifiable one in the sense that an evaluator can design a test of relationships between the systems of input and output variables, and compare the empirical observations with the claims made for the program. This conception, of course, assumes that the program rhetoric accurately reflects the causal model and that the program is properly implemented; that is, all of the claimed causal elements are represented. Unfortunately, most programs vary widely from the ideal, and it is rare, in fact, to find a program with an explicit causal model specifying the relations among input components or between input components and outcomes. Given this latter situation, the evaluator is then faced with several model building tasks which he must perform before the program can receive a fair evaluation.

Formative evaluation activities can play an important role in the development of such a causal model of the program. A formative evaluation can help a program developer to specify the expected outcomes more clearly and to develop appropriate outcome measures, to conceptualize the program, and through the use of program tryouts and corrective feedback to construct the program so that it has a greater likelihood of actually delivering the intended set of input variables of experiences.

When these prior functions have been satisfied, the activities involved in summative program evaluation are less difficult to accomplish. Most summative evaluations, however, are performed on programs which have had little formative evaluation. A school board, curriculum committee, or state agency may decide to adopt a program on a trial basis because the general description of the program appears to meet a generally local or regional need or bias "nice" public relations value. An evaluator is then brought in to gather and summarize data on the program's ability to effect important outcomes. The evaluator is likely to fact multiple problems in attempting to do this evaluation job. The most fundamental problem stems from the likelihood that instead of following a linear development process from a) a theory which specifies that b) certain program components c) will regularly lead to certain outcomes, the program components probably developed first. The outcomes claimed for the program usually followed temporally the development of the actual program components, and the outcomes claimed may not accurately reflect what operationally goes on in the program or what one could logically or empirically expect as an outcome. The rhetoric about the program, or a set of theoretical conceptions about it, is most often the last to occur, as the explanatory and public relations information brochures are made up. Thus when there is the opportunity for this disjunction between the program rhetoric, the operational program components, and the claimed outcomes, the evaluator must decide whether to accept the rhetoric or characterization of the program that is provided by the developers, or to develop an alternative theoretical model based on his own conception of what the actual components of the program are, or to modify or revise the program to fit some other causal model to be tested.

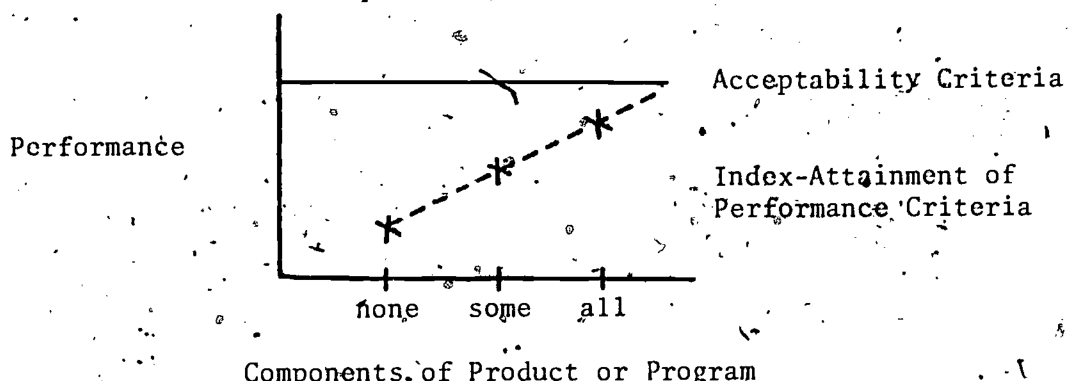
The first alternative is likely to yield a deficient causal model, while the second may lead to a deficient program with an appropriate causal model. Either of these strategies can lead to unmeasured effects, findings of no difference, or both. The third alternative will require that the evaluator work with the developer to construct a more accurate causal model of the program and a more complete and integrated program will be the result. The payoff will come in a fairer test of the program.

Types of Evaluation Studies and Causal Analysis

At least three types of evaluation studies can be distinguished in which causal model analysis might be employed. These are 1) exploratory, 2) confirmatory, and 3) optimization evaluation studies. The three types differ in the bais questions they ask about the program or product being evaluated; and, they differ in the index they employ for examining performance, adjudging program or product adequacy, and arriving at "causal" inferences.

Exploratory evaluation studies have as the basic question, "Is there a program (product)?" That is, have the developers contributed "something" that might be worth continuing to examine, to produce or to promote. The basic research paradigm for examing this question is reflected in the diagram below.

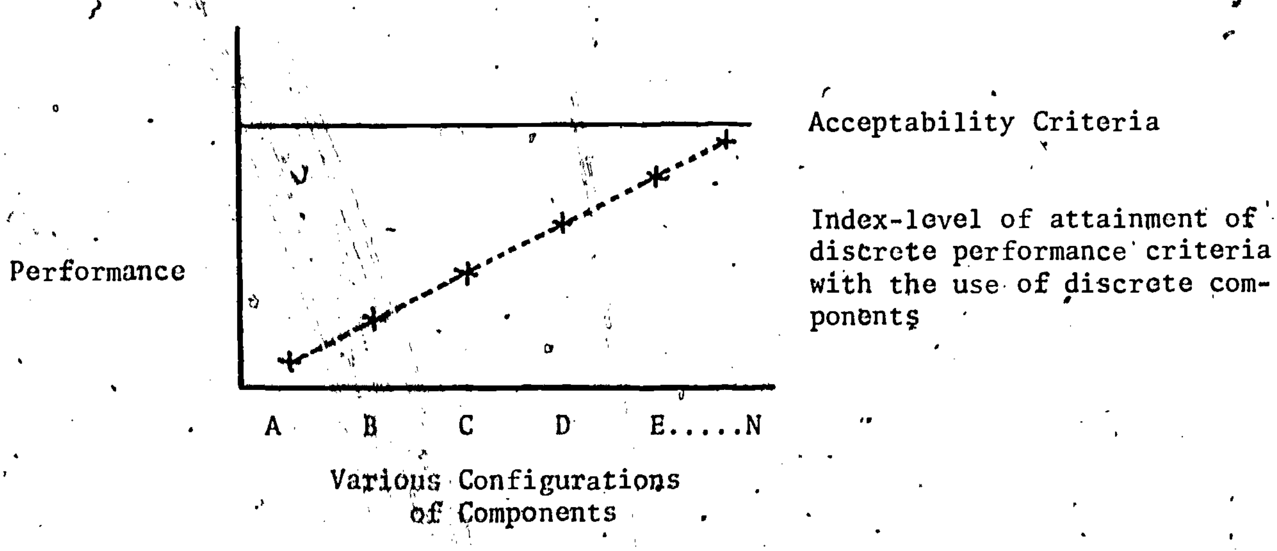
Research Paradigm for
Exploratory Evaluation



Alternatively we could also do exploratory evaluation by asking questions related to, "What kind of performance might I be likely to attain if I used a given set of components?"

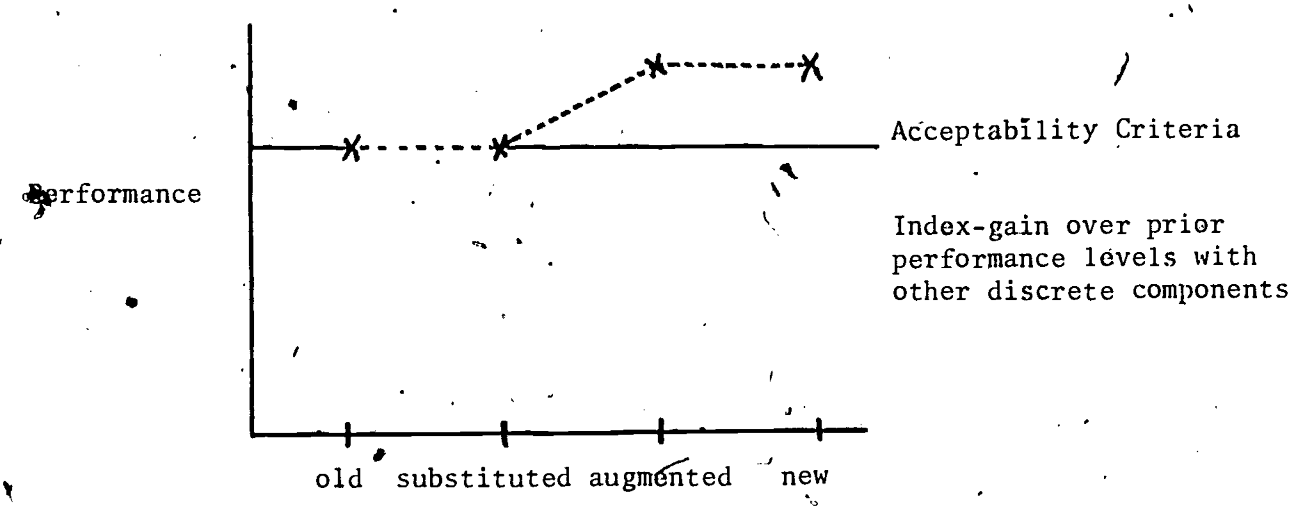
The determination of the existence of some kind of "program" or "product" is inferred from the relationship of the observed performance to the acceptability criteria under at least two levels of implementation ("none" and "some" or "all") of the major components of which the program or product is thought to consist. The hypothetical data in the figure shows how such performance might be plotted to demonstrate a minimal "causal" relationship between the program and the performance. This is the kind of paradigm that is most often used to evaluate the effectiveness of a program or a product. (Alternatively it is also possible to ask, "What kind of performance might I expect to get if I put together a given type of program, based on previously collected data and observation?" However, this type of evaluation is less typically employed.)

Confirmatory evaluation studies have a different basic question underlying their use. The issue is not whether there is some kind of program or product there, but rather, whether the program or product is some specific identifiable subset of program components which are required to generate the program's "effects." The basic research paradigm for examining this question is reflected in the diagram below.



The causal relationship between discrete performance criteria and related components is inferred from the manner in which variations in the level of implementation of discrete program components influence the levels of discrete performances linked logically to them. As the hypothetical data in the figure show, a given configuration of components may be tried to determine if they are a necessary and sufficient condition for the attainment of certain performance criteria. Few evaluation studies have used this paradigm, although evaluators who regularly collect data on the "degree of implementation" of programs are closer to dealing with the appropriate levels of variation issues. They fall short of this possibility of "causal" inferences because too often they fail to connect logically the observed variation of discrete effects to discrete program elements. The latter step is required to have an internally consistent program.

Optimization evaluation is a variation of confirmatory evaluation in which the basic question is now, "Can the program or product be improved?" The basis research paradigm might employ the use of a response surface design as illustrated below.



Configuration of Components

In this type of evaluation the goal may be either the maintenance of performance when substitute components (possibly cheaper or less complicated) are employed, or the maximization of performance by revision or augmentation of existing components, or by the development of new configurations. The "causal" inferences are derived from the continuance or increase of the desired performance level under the substituting or improvement treatment conditions.

Causal Modeling

Much of the work on causal models in non-experimental research represents an attempt to reason causally from correlational findings. The "cause" is inferred from some observed uniformity of relationship between certain sets of variables, derived from a combination of logical and mathematical procedures, and culminates in a set of predictions about what the empirical findings would be if we performed the "experiment" (Blalock and Blalock, 1968; Wittrock and Wiley, 1970; and Goldberger, 1972). As most of the techniques are used, a model is set forth regarding the hypothetical flow of causal influence, and the statistical procedures used attempt to estimate how much change in a dependent variable would be associated with a certain magnitude of change in one or more of the hypothesized causal indicators. The flow of influence may be relatively straight forward, i.e., $x \rightarrow y$ relationship, or it may involve mediated influence as in the case where the x influences y, and y subsequently influences z. It is well known that difficulties may arise in such analyses because of error of measurement, unequal precision of measurement, and sample bias, as well as other factors (Wiley and Wiley, 1970; Hauser and Goldberger, 1971; Cochran, 1972; Cronbach and Furby, 1970; and Wiley, 1973). In addition, where many potential indicators are involved it is sometimes necessary for the research to use additional multivariate techniques to composite and reduce the number of variables into a more manageable set (Wiley, 1970), or to create new pseudo-variates, or to perform various kinds of transformations to normalize distributions and linearize relationships, or to apply more complex solutions (Davis, 1973; Poirier, 1973).

All of the preceding discussion is to point out that there is a somewhat predictable amount of uncertainty in the inferences derived from causal model analysis of non-experimental research data, which is an acceptable or tolerable by-product, because such causal analysis usually provides the researcher with a more precisely explicated model which he may subsequently verify by later formal experimentation. The subsequent verification step is an important one, and it is one of the reasons why these approaches might be referred to as <u>exploratory</u> causal analysis.

Now, if we assume that a causal model is essentially a representation of reality, then when a given causal model is fitted to a set of data there is a test of the veridicality of the particular representation of reality to the observed events. In actual practice the researcher strives to gain credibility for the particular set of explanatory variables which comprises his causal model partly by demonstrating the degree of veridicality of the model, and partly by systematically ruling out other explanatory variables of the same events as Tukey (1954), and later others, have pointed out. One of the more acceptable practices is to attempt to rule out some of these other explanatory variables by the use of statistical designs and sampling procedures which minimize the opportunity for selection bias to occur, or reduce the likelihood that variables not specified or outside of our model can systematically influence the outcomes of interest, or lead to spurious causal attribution.

This would reflect the use of "tight" design features to help arrive at statements of causality within and between program components and program outcomes.

However, suppose on the other hand that we wish to test only one causal model, i.e., some a priori model to which we have a vested interest, and for which all of four to five years of previous instrumentation and technology development work had been allocated, for example, a particular early childhood instructional program with multiple curriculum and teacher training components. The "program" could be considered to be a formalized causal model with lots of identifiable components and potential linkages as represented in the program hardware and software. Such a program if well defined and characterized could be a "strong" causal model in the sense that it, or its set of components, was falsifiable; it could be an "important" model in the sense that it, or its sets of components, would have direct-policy implications for practice; and it could be a "general" model to the extent that it had systematically incorporated, within its formal program components, sets of variables commonly distinguished within other theoretical models related to the same outcomes. Put in another way, when we have performed all of the theoretical extrapolations, the basic and applied research, and completed the design and development of most of the technology for a given educational program, then we are beyond the exploratory stage in causal analysis of program sources of influence and have moved into the confirmatory stage.

In fact, what we literally mean by the use of the term, "program," is that we have developed a system of variables that we have embedded in a set of products and materials. We have a "model" of how this system of variables works, within components and between components, and we believe that each of these components is behaving in a causal fashion as the description of the program states, and in a manner consistent with our theoretical notions. When we have a "program" structured in such detail, then we don't need experimental manipulation of the kind required following conventional exploratory causal analysis. This is because the structure of the program can be directly mapped into a causal model, reflecting its own program-generated manipulations, and verified directly. Moreover, we can be categorically less concerned about spuriousness when there are a large number of predicted causal linkages, since it is not likely that all such linkages could arise accidentally within the mass of detailed data available, and compete with a tightly constructed theoretical rationale and program logic which related the detailed elements of the causal model conceptually. Thus, in effect, we have substituted "tight" theory for "tight" design to guide our causal inferences.

Moreover, when there exists a structured correspondence between the educational program we have developed and the causal model of it, then everything that the components of the program achieve can be achieved by use of the components explicated within the causal model, and can be described in the language of the theoretical framework we have developed for it. Confirmatory causal analysis seeks to determine whether such a structural correspondence exists, and whether the results of the causal model can be systematically obtained from the implementation or use of the formalized "program."

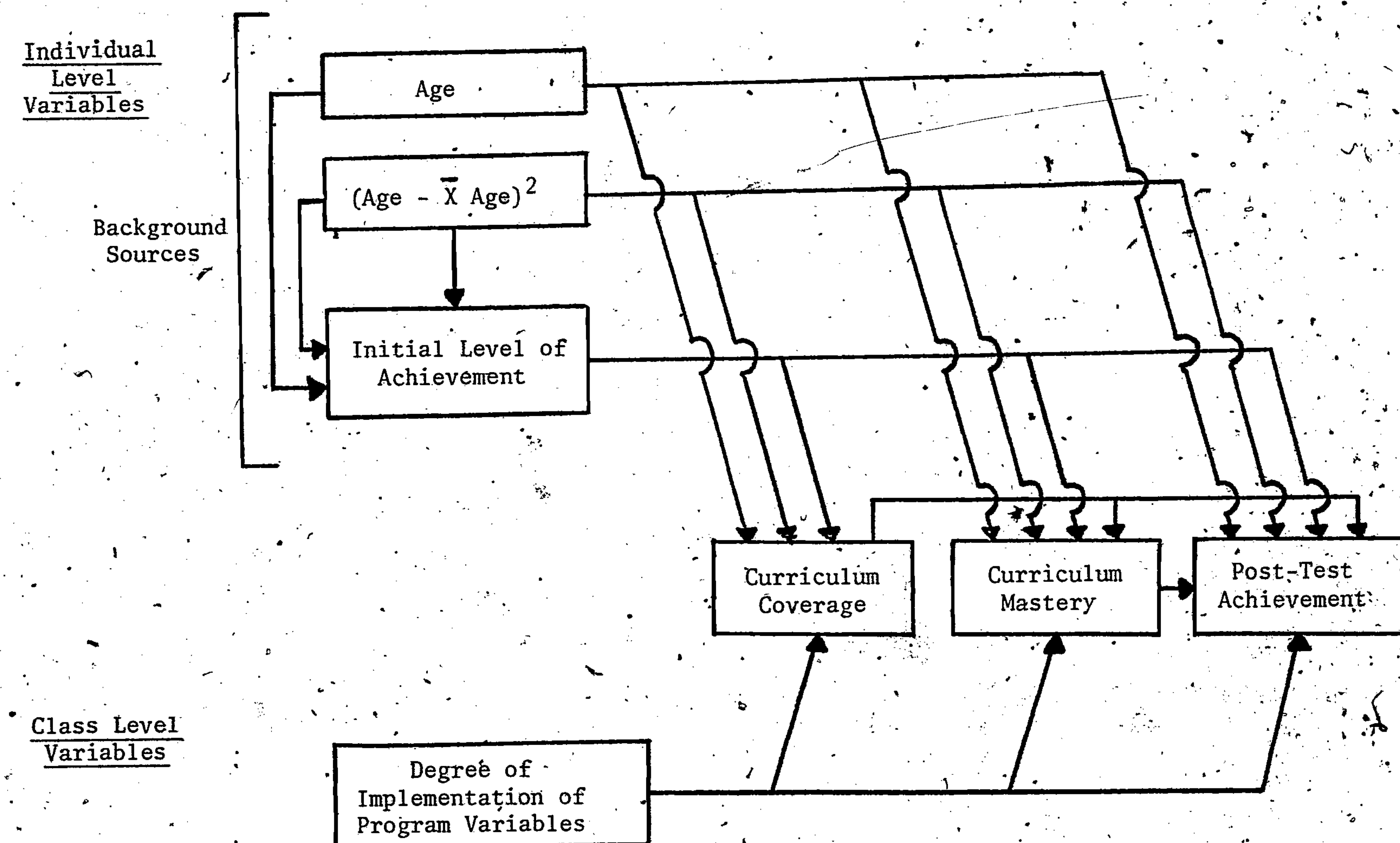
Application of Causal Modeling to the Evaluation of a Program

From 1971 through 1973 a group of evaluators worked closely with the developers of a preschool program. The staffs first worked together to analyze the program, to identify major and minor program components, and to link these components to expected outcomes. This analysis revealed that several important components were either missing or were deficient. The cooperative venture also revealed that the program had no measurement system for tracking a major set of input variables (dealing with teacher's behavior) or assessing the full range of desired outcomes. Once the program was strengthened and the measurement system was created, an evaluation was possible on this program with a theoretical framework in which the system of components of the program were both logically and operationally linked to anticipated outcomes by a common language system within a strong causal model.

The evaluation was conducted during the 1972-73 school year in 54 kindergarten classes in two cities. Degree of implementation measures were collected on the major program input variables and a record of curriculum coverage was kept on each child. Pre and post tests were administered to experimental and control groups on a standardized test, and a test constructed to assess all of the objectives of the program. The original causal model of the program is depicted in Figure 1 (next page). In the model background variables are included which might have had direct effects on program related variables and outcomes, and individual level variables are separated from class-level variables. A two-stage analysis procedure was employed on the data collected during the evaluation effort. In the first stage the effects of age, deviated age, initial level of achievement, curriculum coverage, curriculum mastery, and post-test achievement were tested by a regression analysis with the latter variables serving first as criteria for earlier variables and then as predictors for subsequent variables in the model.

In the second analysis the effects of variation in the degree of implementation of program variables operating at the classroom level were tested using residualized means for initial achievement, coverage, mastery, and post-achievement for the various classes. These means were derived by regressing

Figure 1. Model for Analysis of Probable Causal Influences Derived from Darcee Program

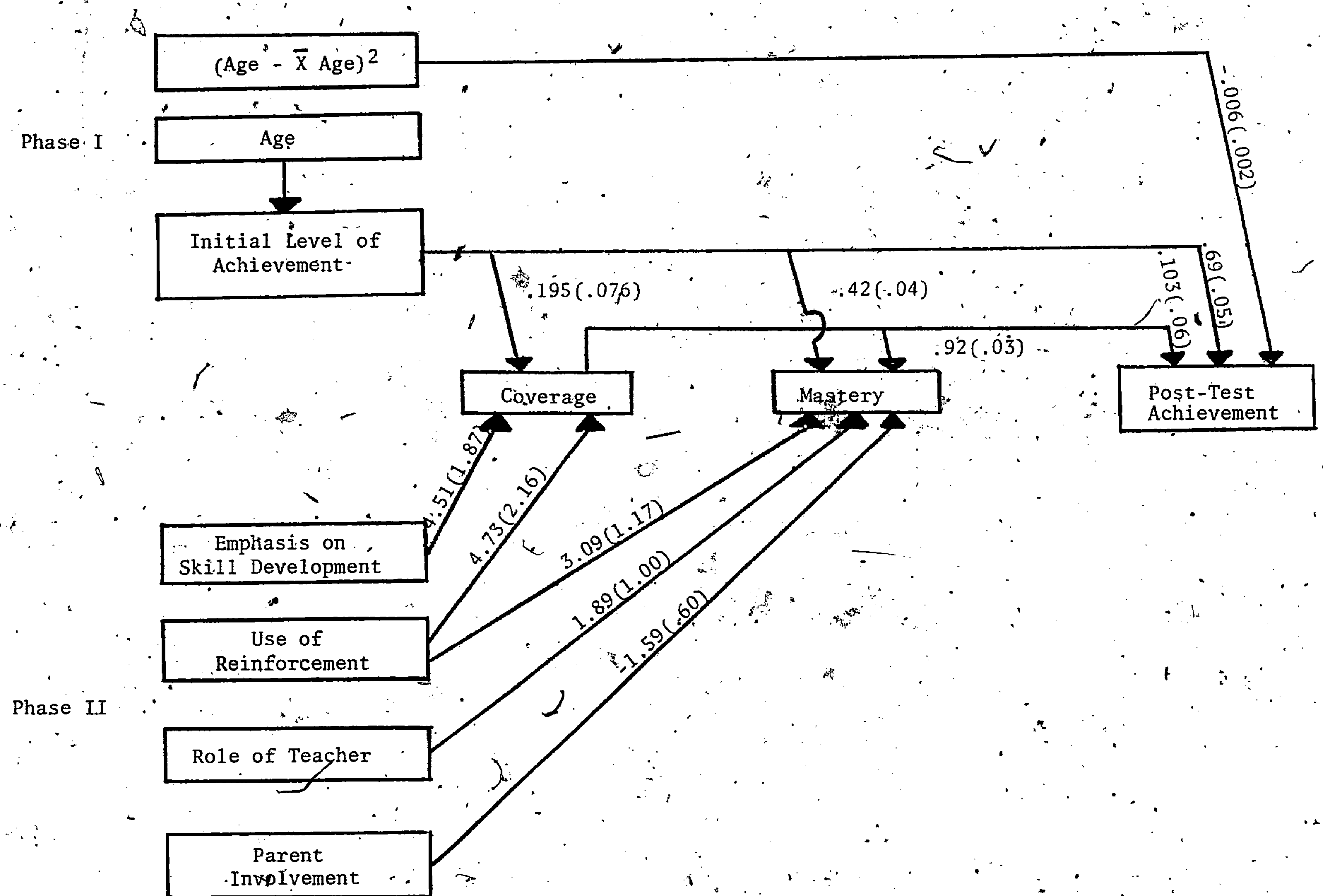


the other five variables on a given variable and then computing the average residual for a given class. The results of the two analyses are shown in Figure 2 (next page). Wherever a program variable has been shown to have a strong relationship to a subsequent program variable or to the major dependent variable, we have presented raw regression coefficients with the standard errors for each in parentheses. The results of the analyses are especially helpful in demonstrating the value of the causal modeling procedures discussed earlier. First, of the eleven "essentials" of the Darcee Program none has been shown to have a direct influence on post-test achievement at the class level, two of the essentials are shown to have any strong direct influence on curriculum coverage, and three essentials have a direct effect on mastery scores. Moreover, for the latter all of the effects on the post-test achievement are mediated through the coverage variable.

Secondly, the initial level of achievement has been shown to be a strong source of influence on coverage, mastery, and post-achievement. In fact, the persuasiveness of the effects of this variable, especially as it influences the curriculum coverage scores, highlights a previously unconsidered issue in educational program evaluation.

The degree of relationship between initial level of achievement and coverage was a strong positive one, and this means that teachers were covering what children had already indicated they could master on the pre-test. The evaluation issue centers on the consequences to be derived from individualization of instruction if the individualization strategy is to work with skills that children already know. An equally important consideration is how program evaluation procedures might detect such consequences. Certainly the use of micro-evaluation techniques is necessary when program influences are likely to be subtle or complex.

Figure 2. Causal Influences Within Program as Derived from Regression Analyses



Since it was not the purpose of this presentation to deal with all of the results from the evaluation of this pre-school program, we have not presented all of the evidence we have collected. The use of causal modeling has provided us with insight into how programs work and how well they work. In the case of this program other evidence clearly indicates that the program was extremely successful in training teachers to behave in a fashion consistent with the objectives of training, i.e., to implement the Darcee "essentials." That these "essentials" were minor sources of influence on achievement may be less reflective of the planned part of the program, than it reveals the non-productive way in which the curriculum activities were actually used to facilitate achievement. The opportunity to focus on this kind of problem and to correct it is at least one of the alternatives available and one of the benefits such types of causal analyses can lead to.

Bibliography

Blalock, H. M., Jr. and Blalock, A. B. (Eds.). Methodology in Social Research. McGraw-Hill Book Co., Inc., New York, 1968.

Cochran, W. G. "Some effects of errors of measurement on linear regression." Proceedings of the Berkeley Symposium on Mathematical Statistics and Probability, Vol. , Berkeley and Los Angeles, University of California Press, 1972.

Cronbach, L. J. and Furby, L. "How should we measure change--or should we?" Psychological Bulletin, 76, 1970, 68-80.

Cronbach, L. J., Glaser, G. C., Nanda, H., and Rajarantnam, H. The Dependability of Behavioral Measurements: Theory and Generalizability for Scores and Profiles. New York, Wiley, 1972.

Davis, J. W. "Multiple regression technique for Pth degree polynomials with and without linear cross products." NASA TN D-7422, December 1973.

Goldberger, A. S. "Structural equation methods in the social sciences." Econometrica, Vol. 40, November 1972, No. 6.

Hauser, R. M., Goldberger, A. S. "The treatment of unobservable variables in path analysis." Chapter 4 in Sociological Methodology, 1971, H. L. Costner (Ed.), Jossey-Bass Publishing Co., 1971.

Poirier, D. J. "Piecewise regression using cubic splines." Journal of the American Statistical Association, September 1973, Vol. 68, No. 343.

Tukey, J. M. "Causation, regression, and path analysis." In Kempthorne, O., et al. (Eds.), Statistics and Mathematics in Biology, Ames, Iowa, Iowa State College Press, 1954.

Wiley, D. E. "Design and analysis of evaluation studies." In Wittrock and Wiley (Eds.), The Evaluation of Instruction: Issues and Problems, New York, Holt, Rinehart and Winston, 1970.

__________. "The identification problem for structural equation models with unmeasured variables." In Goldberger, A. S. and Duncan, O. D. (Eds.), Structural Equation Models in the Social Sciences, New York, Seminar Press, 1972.

__________ and Wiley, J. A. "The estimation of measurement error in panel data." American Sociological Review, 1970, 35, 112-117.

Wittrock, M. C. and Wiley, D. E. (Eds.). The Evaluation of Instruction: Issues and Problems. New York, Holt, Rinehart and Winston, 1970.