DOCUMENT RESUME

ED 045 706 TM 000 275

AUTHOR	Stufflebeam, Daniel L.
TITLE	The Use of Experimental Design in Educational Evaluation.
INSTITUTION	Ohio State Univ., Columbus. Evaluation Center.
PUB DATE	Mar 70
NOTE	14p.; Paper presented at the national convention of the American Educational Research Association, Minneapolis, Minnesota, March 1970
EDRS PRICE	EDRS Price MF-$0.25 HC-$0.80
DESCRIPTORS	Educational Experiments, *Educational Research, Evaluation Criteria, *Evaluation Techniques, Individual Needs, Input Output Analysis, *Models, *Program Evaluation, *Research Design

ABSTRACT

Experimental design has only limited utility inthe field of educational evaluation and, in general, the methodology of the former does not equal the methodology of the latter; in fact, experimental design plays a limited role in the total framework of educational evaluation. Experimental design does have potential utility in the areas of input and product evaluation; however, it has little utility within the areas of context and process evaluation. The utility of experimental design can be increased by following a set of procedures that do not require the use of a common criterion instrument and a uniform decision rule for all students in the experiment. This will allow an investigator to judge a program in terms of the number of students for whom it was successful. A schema for the development of product evaluation designs is included. (Author/CK)

THE USE OF EXPERIMENTAL DESIGN IN EDUCATIONAL EVALUATION

Daniel L. Stufflebeam

The Ohio State University
Evaluation Center

Paper read at the national convention of the
American Educational Research Association
Minneapolis, Minnesota
March 1970

THE USE OF EXPERIMENTAL DESIGN IN EDUCATIONAL EVALUATION

Three purposes have guided the development of this paper. The first is to show that experimental design has only limited utility in the field of educational evaluation. The second is to specify the instances in which experimental design can make important, although limited, contributions to educational evaluation. The third and final purpose is to propose a means by which the utility of experimental design for evaluation can be increased.

Definitions of Evaluation and Experimental Design

Before one can assess the utility of experimental design for evaluation, it is necessary to define what is meant by the terms evaluation and experimental design.

The Phi Delta Kappa National Study Committee on Evaluation has defined evaluation as the process of delineating, obtaining, and providing useful information for judging decision alternatives.

The basis for this definition rests in dictionary definitions of itw two key terms. Among other ways, Webster defines evaluation as the ascertainment of value and defines decision as the act of making up one's mind. The need to make up one's mind connotes the existence of competing alternatives. In order to choose one alternative over the other(s), their relative values must be ascertained. Hence evaluation may be defined as the process of ascertaining the relative values of competing alternatives.

Simply stated, evaluation is the process of providing information for decision-making.

Since the purpose of evaluation is to provide information for decision-making, the decisions to be served must be known. Generally, these decisions may be divided into four classes called planning, structuring, implementing, and recycling decisions. Planning decisions pertain to the selection of objectives. Structuring decisions are those involved in designing projects to achieve stated objectives. Those required for operationalizing and executing a project design are referred to as implementing decisions, and recycling decisions refer especially to the judgment of and reaction to project results.

Since there are four kinds of decisions to be served, there are also four kinds of evaluation. Context evaluation serves planning decisions by identifying unmet needs, unused opportunities, and underlying problems which prevent the meeting of needs or the use of opportunities; input evaluation serves structuring decisions by projecting and analyzing alternative procedural designs; process evaluation serves implementing decisions by monitoring project operations; and product evaluation serves recycling decisions by determining the degree to which objectives have been achieved and by determining the cause of the obtained results.

Given this definition of and rationale for evaluation, let us next define experimental design.

Traditionally, experimental design has been the recommended strategy for determining the effectiveness of projects. A group of subjects is chosen and randomly divided into two subgroups; one group is assigned to an experimental treatment (for example, modern mathematics), and the other is assigned to a control condition. The two conditions are then imposed and the subjects are measured on a common criterion instrument at the end of the experiment. Analysis of the different performance levels for the two groups then provides causal statements about the differential effectiveness of the competing conditions.

The Role of Experimental Design in Evaluation

With the above definitions of evaluation and experimental design it has been possible to perform logical analyses of the utility of experimental design for each of the four evaluation types that were specified. The results of these analyses are contained in Tables 1 through 4, which pertain respectively to context, input, process, and product evaluation.

Each table contains four columns. The first includes a list of questions which are illustrative of those that should be answered by the type of evaluation that is pertinent for that table. Column 2 identifies the kind of information that is needed to answer each question. Column 3 contains judgments of the relevance of experimental design for obtaining the specified information. Column 4 lists alternative techniques which are judged to be equal or superior to experimental design in obtaining the specified information.

An examination of the four tables quickly reveals that experimental design is judged to have much relevance for product evaluation, minor relevance for input evaluation, and no relevance for context and process evaluation.

TABLE 1

A Logical Analysis of the Utility of Experimental Design for Context Evaluation Studies

Illustrative Questions	Illustrative Information Requirements	Relevance of Experimental Design	Illustrative Alternative Techniques
What unmet needs exist in the context served by a particular institution?	System goals, system performance, and the discrepancy between the two.	None	System analysis, Management information system.
What improvement-oriented objectives should be pursued in order to meet identified needs?	Diagnoses of the problems which account for discrepancies between system goals and system performance.	None	Review of the literature, Case studies.
What improvement-oriented objectives will receive the endorsement and support of the community?	Descriptions and analyses of community values regarding possible objectives that could be sought with program improvement resources.	None	Sample surveys, Community television forum.
Which of a set of objectives are most feasible to achieve?	Estimates of the technological tractability of possible objectives that could be sought.	None	Review of the literature, Consultation with a panel of experts.

TABLE 2

A Logical Analysis of the Utility of Experimental Design
for
Input Evaluation Studies

Illustrative Questions	Illustrative Information Requirements	Relevance of Experimental Design	Illustrative Alternative Techniques
Does a given project strategy provide a logical response to a set of specified objectives?	Statements of expert judgment.	None	Proposal reviews by panels of experts.
Is a given strategy legal?	Legal opinion.	None	Legal counsel.
What strategies already exist with potential relevance for meeting the established objectives?	Identification and analysis of strategies that are already operating in similar institutions, or that are being developed in research and development institutions.	None	Use of ERIC, Visitations to other institutions and to R. and D. agencies.
What specific procedures and time schedule will be needed to implement a given strategy?	Identification of project events and activities, development of a network to show the interrelationships between events and activities, and assignments of time estimates to the activities.	None	PERT, CPM.
What are the operating characteristics and effects of competing strategies under pilot conditions?	Comparative data pertaining to the costs and benefits of competing strategies.	Strong, if the expense can be justified.	Querying ERIC, Visitations to sites where the competing strategies are operating.

TABLE 3

A Logical Analysis of the Utility of Experimental Design for Process Evaluation Studies

Illustrative Questions	Illustrative Information Requirements	Relevance of Experimental Design	Illustrative Alternative Techniques
Is the project on schedule?	Comparison of actual and scheduled completion dates for project events already completed.	None	PERT, CPM.
Should the staff be retrained or reoriented?	Report concerning the extent to which staff understand their roles, are motivated to perform them, and actually are doing so.	None	Classroom observation, Interviews, Unobtrusive measures such as the amount of coffee consumption.
Are the facilities and materials being used adequately and appropriately?	Report concerning the extent to which materials and facilities are being used in the prescribed manners and amounts.	None	Classroom observation, Interviews, Inventory of materials and facilities use.
What major procedural barriers need to be overcome?	Report representing the perceptions of the project staff and participants concerning the barriers that they think exist and should be overcome.	None	Interviews, Suggestion box, Forums for the discussion of this issue.

TABLE 4

A Logical Analysis of the Utility of Experimental Design for Product Evaluation Studies

Illustrative Questions	Illustrative Information Requirements	Relevance of Experimental Design	Illustrative Alternative Techniques
Are objectives being achieved?	Comparison of attainment measures with objectives or with the performance of a control group.	Strong	Comparison of attainment measures with absolute standards.
What probabilistic statements can be made about the relationship between procedural specifications and actual project attainments?	Inference about the causal relationship between means and outcome data.	Strong	None as strong as experimental design.
To what extent were the varied needs of individual students met as a result of the project?	The number of successes occurring for individual students in the program, in terms of their individual needs.	Weak	Case studies of a random sample of cases and non-parametric analysis of the results.
What is the long-range worth of the actual achievements in relation to the mission of the host institution?	Cost/benefit projections under the assumption that the program being tested would be installed.	Weak	Cost/benefit analysis.

Thus, the methodology of evaluation is not equal to the methodology of experimental design. Neither should experimental design be dismissed as entirely irrelevant to the field of evaluation. Rather it should be recognized that experimental design should occasionally be utilized in input evaluation and that it has a major role to play in product evaluation.

A Paradigm Designed to Increase the Utility of Experimental Design in Educational Evaluation

If the assumptions required by experimental design can be met, the evaluator has a powerful and efficient tool for answering certain input and product evaluation questions. By its use, relatively unequivocal statements can be made that a program was or was not more effective than a competing program in producing a desired effect. In the final analysis, this is the type of information that decision-makers and those they serve want.

However, several problems block the effective use of experimental design. Usually, assumptions of constancy of treatment across both subjects and time and additivity of effects cannot be met. As any teacher knows. different students require different treatments and different students learn at different rates. Also, and perhaps as a consequence of violating too many assumptions, the use of experimental design in education has uncovered few significant differences between experimental and control conditions. Then, if experimental design doesn't perform well in practice, why not, and what can be done to obtain the needed information for explanation of outcomes?

It is proposed here that the general principles of experimental design can and should be employed in certain kinds of input and product evaluation situations. What is needed is variation in how the experimental or quasi-experimental design is applied. Specifically, the requirement that each child receive the same treatment and that the same definition of success apply to all children must be eliminated. Johnny may like arithmetic but have a problem learning to count to 20, and Mary sitting next to him may be developing an understanding of the concept of multiplication but have a very negative attitude toward arithmetic. Johnny and Mary would not need identical instructional treatments. Furthermore, success for the two at any given point in time would not be the same. One would require a test of addition, while the other would require a test measuring attitude toward arithmetic as well as a test in multiplication. In effect, does the experimental program successfully meet the individual needs of the students served? If individualization is a valid concept, this question seldom can be answered by using a common criterion instrument and a uniform decision rule. Adherence to these two rules often has been the downfall of the use of experimental design in the past. If the definition of success varies for the individuals in a program, then any evaluative effort employing only one standard is doomed to failure. Much significant information is not collected, and much of that which is collected is washed out through averaging and interpretation against a single criterion measure and a single standard. How, then, can this dilemma be overcome?

One answer seems simple. Figure 1 summarizes this answer. In effect, Figure 1 is a paradigm designed to expand the utility of experimental design for evaluation.

Starting at the left, the first three columns illustrate that a sound experimental design could be selected, that a sample could be randomly selected from a specified population, and that experimental and control groups could be assigned randomly, as in the usual case. Objectives could next be assigned for each child based upon the context information about him, as shown in column 4.

Then, as shown in column 5, within the constraints of the experimental and control conditions, input evaluation could be used to assign the treatments that are most relevant for each child.

Next, as illustrated by columns 6 and 7, product measures and standards of success could be specified in light of the objectives assigned to each child. At the end of the project cycle, the specified criterion instruments could be administered to each child, and the obtain measurements could be classified in nominal terms such as success or failure. The results from the experimental and control groups could then be compared, in accordance with the decision rule specified in column 8 and by use of a nonparametric test statistic such as chi square specified in column 9. This would allow the investigator to state unequivocally that the results from program A were or were not superior to those from program B in serving the varied needs of pupils from the specified population.

FIGURE 1

SCHEMA FOR THE DEVELOPMENT OF PRODUCT EVALUATION DESIGNS

1. Basic Design	2. Basic Population of Interest	3. Random Sample $n_d + n'_d$ Subjects	4. Objectives	5. Treatments	6. Criterion Instruments	7. Success Standards	8. Decision Rule	9. Test Statistic	10. Interpretation
(Specification of a pre-, quasi-, or true experimental design.)		E_1	O_1	T_1	C_1	S_1			
		E_2	O_2	T_2	C_2	S_2			
		E_3	O_3	T_3	C_3	S_3			
		. . .	. . .	. . .	. . .	. . .			
		E_d	O_d	T_d	C_d	S_d			
Example:							Example: $H_1: n_s = n_s'$	A Example:	(Conclusions as to the relationship between treatments and effects.)
R X O	N						$H_2: n_s \neq n_s'$	Chi Square	
R O							α: .05	α: ____	
		E'_1	O'_1		C'_1	S'_1			
		E'_2	O'_2		C'_2	S'_2			
		E'_3	O'_3		C'_3	S'_3			
		. . .	. . .		. . .	. . .			
		E'_d	O'_d		C'_d	S'_d			

Columns 2–4: Based upon context evaluation data

Column 5: Based upon input evaluation data

Columns 6–9: product evaluation

Column 10: Comparison of process & product data

These results could be interpreted further by reference to process evaluation information which describes the experimental and control conditions as they actually occurred. For example, if there were no significant differences between experimental and control effects, it would be important to know whether the treatment and control conditions had actually been applied as intended.

The significant aspect of this strategy is that it considers all instances of success or failure in terms relevant to each instance. In one sense, this is like mixing apples and oranges. But the point is to convert the ordinal and interval data about each child to nominal data, which can be grouped and analyzed for the program as a whole. This general strategy can be applied to virtually all known pre-experimental, experimental, and quasi-experimental designs. Thus, this simple set of process steps should extend the utility of experimental design in input and product evaluation studies.

Conclusion

In this paper I have attempted to make three points.

First, the methodology of educational evaluation is not equal to the methodology of experimental design; in fact, experimental design has a very limited role to play within the total framework of educational evaluation.

Second, experimental design does have potential utility in the areas of input and product evaluation. However, it appears to have no utility within the areas of context and process evaluation.

Third, the utility of experimental design can be increased by following a set of procedures that do not require the use of a common criterion instrument and a uniform decision rule for all students in the experiment. In effect, this will allow an investigator to judge a program in terms of the number of students for whom it was successful.