

The Journal of Educational Evaluation
evaluation comment

Center for the Study of Evaluation
UCLA

THEORY-BASED EVALUATION

Carol Taylor Fitz-Gibbon
Lynn Lyons Morris

What do we mean by theory-based evaluation?

A theory-based evaluation of a program is one in which the selection of program features to evaluate is determined by an explicit conceptualization of the program in terms of a theory, a theory which attempts to explain how the program produces the desired effects. The theory might be psychological, such as a theory of child development (e.g. Piaget's) or a theory of learning (e.g. $S \rightarrow R$ theory), or social psychological (e.g. attitude change theories; organization theories) or philosophical (e.g. the "Summerhill" philosophy). The essential characteristic is that the theory points out a causal relationship between a process A and an outcome B. $A \rightarrow B$; that is, A leads to, or causes, B. (A, of course, may consist of many necessary components or stages – the whole process deemed necessary, by the theory, to produce B.) Thus, by a "theory-based" evaluation, we do not mean an evaluation based on a theory about evaluation. We mean rather, one based on a theory about how a program operates.

- writing on theory-based planning and evaluation

Carol Taylor Fitz-Gibbon is a Senior Research Associate at the Center for the Study of Evaluation and Co-Director of the CSE Summative Evaluation KIT

Lynn Lyons Morris is a Senior Research Associate at the Center for the Study of Evaluation and Co-Director of the CSE Summative Evaluation KIT

Adrienne Bank is a Senior Research Associate at the Center for the Study of Evaluation and Director of the CSE Program Planning KIT

The KITs mentioned above are part of the CSE Elementary School Evaluation KIT Series, which includes separate KITs on Needs Assessment, Program Planning, Formative Evaluation, and Summative Evaluation. Each contains materials designed to enable elementary school or district personnel to deal with the evaluation area treated by the particular KIT.

The two articles in this issue were initially stimulated by KIT development activities at CSE. In addition, a reading of Rodney Skager's evaluation of an alternative school led to the coining of the term "theory-based" evaluation.

In the first article, Fitz-Gibbon and Morris describe what they mean by theory-based evaluation and discuss some broad applications of this kind of evaluation. They illustrate the differences between theory-based and current evaluation genres, stressing the usefulness of the former in terms of selecting the most relevant program features for evaluation.

In the second article, Bank discusses a specific application of theory-based evaluation, focusing on teaching models as alluded to in the first article. Bank reviews the seven models presented in the Program Planning KIT which form a strategy enabling school personnel to plan and then evaluate a program in accordance with a particular theory or philosophy of instruction.

Both articles present some of the problems and advantages implied by theory-based evaluation.

Perhaps the inclusion of philosophies as theories needs some justification, especially as we shall subsume "models" (see Joyce & Weil, 1972) under the same rubric. We justify the inclusion by defining a philosophy as a set of attitudes. There has been extensive consideration of the nature of attitudes in the literature of social psychology and those models of which we are aware assert that attitudes have cognitive and affective components. The affect-cognition model (Rosenberg, 1960) is presented as an example since other models seem readily reducible to it. Rosenberg postulates that a person's attitude towards X is a function of his expectation (subjective probability) that X leads to Y and his emotional response to Y.

[Attitude to X] = f [subjective probability X→Y, emotion associated with Y]

In the cognitive component, there is a postulated causal relationship: the perceived probability that X→Y. Attitudes involve theories, therefore, since they postulate causal relationships. For this reason philosophies, like theories, suggest variables to study and imply that certain outcomes are likely to occur from certain processes.

The term "theory-based" evaluation, then, means an evaluation based on a model, theory, or philosophy about how the program works; a model, theory, or philosophy which indicates the causal relationships *supposedly* operating in the program.

How does theory-based evaluation differ from the kind of evaluations currently produced?

As indicated by the preceding definition of theory-based evaluation, the major impact of choosing a theory-based mode of evaluation occurs at that every early stage of evaluation when a selection is made of the program features or variables which will be studied. One cannot measure or observe or report on everything about a program: inevitably, one selects. Does one choose to measure atomistic variables such as positive words per minute uttered by the teacher, concrete "variables" such as materials used, cognitive variables such as reading difficulty of materials, molar variables such as teacher "directiveness", monetary variables such as cost per child, social-psychological variables such as peer acceptance and roles enacted, sociological

variables such as socioeconomic status of teachers and students, or such a prosaic but perhaps powerful variable as hours spent on instruction? The possibilities are endless.

Stake (1967, p. 536) has suggested that deciding which variables to study is an "essentially subjective commitment" in evaluation. Such a contention has the effect of closing discussion on the matter, relegating the problem to some intractable region of pure opinion. We do not agree with this position. The choice of variables to study need not remain a matter of opinion. Eventually, we will want to study those variables which explain the most variance in the outcomes of interest. Hopefully some of these variables will be manipulable and yield powerful positive results in improving educational practice. Discovering what these variables are is a major task for educational research, and not one that we can assume has already been accomplished.

When a theory-based evaluation is planned, the variables selected for study are those which a theory (this might be, but does not necessarily have to be, the theory on which the program itself is based) indicates are crucial in producing the desired program outcomes. The theory chosen is stated and the degree of its operationalization within the program is documented. The methods of theory-based evaluation would resemble, then, those traditionally used to measure degree of program implementation: questionnaires and interviews, examination of records and, particularly, classroom observations.

An example of a theory-based evaluation

Skager's evaluation of the Los Angeles Alternative School (Skager, Morehouse, Russock, & Schumacher, 1973) struck us as a qualitatively different kind of evaluation, an approach to evaluation meriting a term of its own to describe its methodology. Skager and his associates were confronted with the task of evaluating an "alternative school" which was in its first year of operation. What variables were chosen for study? Skager based the selection of variables on a model and a philosophy. The model was Carroll's (1963) model of the school learning process from which Skager deduced, for example, the need to contrast the instruction provided to students who were high and low in "perseverance". The philosophy which informed the selection of variables was Neill's (1969) "Summerhill" philosophy, a cornerstone of which was understood to be that students should choose freely what and when they studied, if indeed they study at all. Working from these "theories" (Carroll's model and Neill's philosophy), Skager and his co-workers selected the variables which the evaluation would measure. Using Carroll's model, the extent to which low and high persevering students were treated differently in the school was assessed and taken as one indication of the quality of instruction. Selecting a crucial variable from Neill's philosophy, the manner in which students selected their activities was investigated, permitting a well-supported statement to be made that "the notion of freedom of choice for the individual child (was) a reality" (Skager *et al.*, 1973, p. 42). This indicated a successful operationalization of Neill's philosophy.

We can note immediately several reasons why this theory-based evaluation was highly appropriate: (a) The program being evaluated was based to some extent on Neill's philosophy. In a sense, the theory-based evaluation was necessary as a check

on program implementation, if for no other reason. (b) The audience for the evaluation report wanted to know if the philosophy had been operationalized. Many parents may have wanted the philosophy to be implemented out of a here-and-now quality-of-life concern, being willing to trust that the ultimate outcomes would be desirable, as predicted by the philosophy. (c) Many of the *outcomes* of greatest concern were not precisely measurable. One may find evaluators who will agree to measure creativity, mental health, and self-esteem, but one should not place great trust in such measurements. The existence and measurability of some variables have yet to be established. (d) Many important desired outcomes were of the nature of long-term consequences rather than immediately measurable outcomes. That is to say, even if over the long term some measures of mental health and vigor could be obtained, this does not mean that beneficial effects of a program could be detected at the end of one year.

When should theory-based evaluation be used?

We shall now attempt to typify situations in which theory-based evaluations appear to be a desirable, perhaps necessary manner of operation. Briefly, these are situations in which evaluation of *process* (A, in the $A \rightarrow B$ conceptualization described earlier) must be attempted either because we are evaluating an instructional program that is itself based on theory or because the client asks questions to which only theory-based answers are now possible. Each situation is discussed below.

Theory-based evaluation should be done when the program to be evaluated is itself based upon a theory, model, or philosophy. An example is provided by the kind of program which would result from the use of the Teaching models component of the CSE Program Planning KIT. School personnel using these materials plan a program in which teaching is structured to conform to models which rest on theories or philosophies about people and schools. An evaluator called upon to evaluate such a program should identify the operational components of the model and develop ways to measure these.

Theory-based curriculum depends on a functional mapping of theory into practice, and practice to outcomes. Theory-based evaluation, in this case, expands the evaluator's job to assessing "goodness of fit" of at least three factors:

- the fit between the theoretical interpretation of the model by program personnel and the evaluator's interpretation (which should represent an informed one)
- the fit between each of these interpretations and the operationalization of the program
- the fit between outcomes predicted by the model and observed outcomes (if the model indeed predicts outcomes that are measurable)

Theory-based evaluation should be employed where an instructional program aims toward distant or tangible outcomes. Outcomes might be unobservable at a given point in time if they involve some slow change, such as character development, attitude change, or the development of complex problem-solving competencies. Such outcomes not only present enormous problems for measurement, but they also cannot be expected to reach a measurable magnitude in a year or two. They are outcomes

that are intangible or so remote in future time that their effects might not be observable within the short time span of most evaluations.

Remoteness and inconcreteness of objectives seem, in fact, to be particularly characteristic of the humanistic trend in education. It seems that it will be good practice for the educational evaluator to adapt to this. After all, schools cannot and should not be forbidden to concern themselves with such complex, but important concepts as a child's self-esteem, and yet evaluation of a program might be required before such outcomes become evidence or measurable. The evaluator should therefore expand his or her repertoire to include not only evaluation based on the measurement of immediate outcomes, but also theory-based evaluation, evaluation which recognizes and holds the program staff accountable for implementing the theory adopted for the instructional program. When the outcome B is postulated, *by a theory*, to follow from a certain process, A, then the *process* can be evaluated *prior to the time when outcomes might possibly be measured*. In these cases, the job of the evaluator, it seems, alters in focus: The evaluation question becomes

Have the variables which theory indicates are crucial to the program actually been operationalized?

In the absence of a guiding theory to follow, the evaluator can only consider immediate and measurable outcomes. Such an approach might be applauded as hard-nosed realism, but we believe that a concern with long-term consequences can only be predicted by making use of theories, of "process A causes (eventually) outcome B" postulations. We agree that the social sciences do not currently provide us with theories that are as reliable as those in the natural sciences, but there seems to be enough theoretical richness in psychology and sociology, for instance, to supply the evaluator with at least a place to begin.

A couple of cautions

One caution which needs to be advanced regarding theory-based evaluations is that citing unproven theory might inadvertently become an "appeal to authority" when included in reports to lay audiences. When evaluations of the kind we are discussing are performed, their theory-based nature should be carefully explained; the evaluator must stress the "if this theory is correct" basis of his or her conclusions. A second problem we foresee is that theory-based evaluation will require quite broadly educated and informed evaluators. The conceptualizations and operationalizations demanded will require greater familiarity with diverse theories and greater flexibility in designing the evaluation than a simple input/output analysis.

We shall conclude by specifying three benefits of theory-based evaluation.

What advantages can be derived from conducting theory-based evaluations?

We can accept the fact that evaluations are generally conducted to serve immediate decision needs. However, representing as they do an enormous investment of time and money, it is to be hoped that evaluations can also contribute to a growing body of knowledge, and to an understanding of the educational process and of social science in general. Theory-based evaluations carry exactly this kind of bonus. By their

nature, they demand clear statements of the assumptions underlying the choice of variables to study in the evaluation. Furthermore they provide a means for comparing data across evaluations. Most importantly, they provide a means for advancing social science in general, via theory testing.

Exposure of assumptions. At present, evaluations are preponderantly atheoretical (Alkin, Kosecoff, Fitz-Gibbon, & Seligman, 1974). Rarely does the evaluator present any rationale for the choice of variables (a sin of omission almost certainly attributable to tradition rather than to deliberate negligence). A hallmark of theory-based evaluation would be the presentation of a rationale for the choice of variables to study. Of course, demanding that this rationale be explicit exposes the evaluator to criticism which might otherwise be avoided had he or she stuck to tradition and slurred over this step. The selection of variables *does* occur, however, whether it be by subjective hunches, by simple adherence to tradition, or by theory-guided considerations. The requirement to spell out assumptions, while irksome, is frequently salutary in its effects.

Comparability. A set of evaluations which all dealt with diverse programs, but had all examined these programs by evaluating them against, for example, behavioristic learning theory (and therefore studied the same variables), would enable cross-site comparisons to be made in a way that is impossible given the atheoretical, idiosyncratic evaluations of the current genre. In sites which had employed both an experimental design (true control group) and theory-based evaluation, an evaluator might, for example, look at the extent to which the success of the experimental group hinged upon the manipulation of those variables identified by the theory as relevant to learning. If no relationship existed, one recommendation would be to build programs in the future on more predictive theories. While this first treatment of theory-based evaluation is speculative and ignores some complexities, we do believe that theory-based evaluation could advance both knowledge and practice. We might find we could agree with Dewey that theory is in the end the most practical of all things. Indeed, Cronbach (1963) in his highly influential article urging formative evaluation, put forward a similar hope (he uses the term "course" where we have been using the term "program"):

Insofar as possible, evaluation should be used to understand how the course produces its effects and what parameters influence its effectiveness Hopefully, evaluation studies will go beyond reporting on this or that course and help us understand educational learning. (p. 675)

Advancing knowledge. Much that could be said under the rubric "advancing knowledge" has been argued above, but we wish to emphasize here a particular point. The touchstone of a theory is its power to predict. Who among us can predict student learning, for example, solely from observation of the process of schooling? The literature is replete with prediction studies, but (apart from teacher effectiveness studies) they largely concern the prediction of student achievement from prior measurements of *students*. It would be an interesting challenge to competing theories of the process of learning to have several evaluators perform theory-based evaluations of several different programs, with each evaluation team basing the evaluations on a particular theory of learning or instruction. How much evidence could be mustered from these theories which would accurately predict differences in the learning

outcomes of the various programs? If a theory – when used to examine process in the way we have been advocating – could consistently discriminate between effective and ineffective programs, it would provide a solid foundation on which to base both the planning and the evaluation of programs.

If evaluation which advances knowledge seems more like research than evaluation, it should be borne in mind that evaluation is concerned with providing information for decision makers. There is a level of decision making where social planning policy is made and there the decision maker needs to know, not whether Sleepy Gulch has a better math program than Thundertown, but rather what guiding concepts, what theories, work in many situations. The social planning level decision maker needs the comparable, generalizable data of theory-based evaluations.

We advocate theory-based evaluation as one possible mode for evaluation because it seems that through it a number of relationships between educational process and outcomes could eventually be established, and those theories holding the most practical promise could be identified. It may turn out, for example, that even though tinkering with stimulus-response variables reliably produces effects, the magnitude of these effects is small when compared with those obtained from a retooling of the whole social-psychological atmosphere of the classroom. Theory-based evaluation could aid us in determining an effective level of analysis for defining instructional variables, and thereby add to the accretion of a firm core of effective theories upon which the design of instruction could be based.

It should be noted that we do not suggest that theory-based evaluation should become the *only* evaluation approach, but rather that it should be instated as one viable, acceptable evaluation alternative. If the social sciences offered firmly established theories directly applicable to the educational process, then theory-based evaluation would certainly be, from a pragmatic point of view, the *preferred* mode. We realize that the social sciences are far from being able to provide a well-established basis for doing evaluations and that many evaluations will have to rely on the standard input/output measurement approach. We are suggesting, though, that wider use of theory-based evaluation is both a practical immediate need and a promising means of advancing knowledge.

REFERENCES

Alkin, M.C., Kosecoff, J., Fitz-Gibbon, C.T., & Seligman, R. (1974) *Evaluation and decision making: The Title VII experience*. CSE Monograph Series in Evaluation. No. 4. Los Angeles: Center for the Study of Evaluation, University of California.

Carroll, J.B. (1963) A model of school learning. *Teachers College Record*. 64(8): 723-733.

Cronbach, L.J. (1963) Course improvement through evaluation. *Teachers College Record*. 64(8): 672-683.

Joyce, B.R., & Weil, M. (1972) *Models of teaching*. Englewood Cliffs, New Jersey: Prentice-Hall Inc.

Neill, A.S. (1969) *Summerhill*. New York: Hart Publishing Co.

Rosenberg, M.J. (1960) Cognitive reorganization in response to the hypnotic reversal of attitudinal affect. *Journal of Personality*. 28(1): 39-63. Quoted in Jones, E.E., & Gerard, H.B. (1967) *Foundations of social psychology*. New York: Wiley.

Skager, R., Morehouse, K., Russock, R., & Schumacher, E. (1973) Evaluation of the Los Angeles Alternative School: A report to the Board of Education of the Los Angeles Unified School District. Los Angeles: Center for the Study of Evaluation, University of California.

Stake, R.E. (1967) The countenance of educational evaluation. *Teachers College Record*. 68(7): 523-540.

CENTER FOR THE STUDY OF EVALUATION

UCLA

Marvin C. Alkin
Director

EVALUATION COMMENT, the Journal of Educational Evaluation, is published by the Center for the Study of Evaluation, University of California, Los Angeles. Founded in June, 1966, by the U.S. Office of Education under the Cooperative Research Act, CSE is one of eight university-based centers and eleven regional laboratories conducting educational research and development for the National Institute of Education. It is the only federally funded research and development center working exclusively on problems in educational evaluation.

Each issue of the journal will discuss topics in educational evaluation by presenting article on evaluation theory, procedures, methodologies, or practices. For each issue of the journal, CSE will request contributions on a specific topic in evaluation from recognized experts in the field; unsolicited manuscripts will not be accepted.

EVALUATION COMMENT has a current circulation of over 10,000 readers. Each scholar, researcher, or practitioner on our mailing list receives a free copy of the journal. Where additional copies are needed readers are encouraged to reproduce the COMMENT themselves. To be placed on our mailing list, please write to:

James Burry, Managing Editor
EVALUATION COMMENT
Center for the Study of Evaluation
145 Moore Hall
University of California, Los Angeles
Los Angeles, California 90024

Editorial Board

Marvin C. Alkin	Chairman
John Evans	Office of Planning, Budgeting, and Evaluation, United States Office of Education
William Johnston	Superintendent, Los Angeles Unified School District
John B. Peper	Assistant Superintendent for Planning, Milwaukee Public Schools
Robert E. Stake	Associate Director of the Center for Instructional Research and Curriculum Evaluation, and Professor of Educational Psychology at the University of Illinois
Daniel L. Stufflebeam	Professor of Educational Leadership and Director of the Evaluation Center, Western Michigan University