DOCUMENT RESUME

ED 050 157 TM 000 549

AUTHOR	Berlak, Harold
TITLE	Naturalistic Observation as a Research Instrument in Curriculum Development.
INSTITUTION	Washington Univ., St. Louis, Mo. Graduate Inst. of Education.
PUB DATE	Feb 71
NOTE	16p.; From symposium "Participant Observation and Curriculum: Research and Evaluation" presented at the Annual Meeting of AERA, New York, N.Y., Feb 1971
EDRS PRICE	EDRS Price MF-$0.65 HC-$3.29
DESCRIPTORS	*Behavioral Science Research, *Classroom Observation Techniques, Curriculum Development, Curriculum Evaluation, *Curriculum Research, Data Collection, Decision Making, Educational Practice, Educational Theories, *Evaluation Methods, Formative Evaluation, Interaction Process Analysis, Program Effectiveness, *Research Methodology, Research Problems, Research Tools
IDENTIFIERS	*Naturalistic Observation

ABSTRACT

A number of problems are identified and questions raised about the usefulness of conventional instruments of educational and psychological measurement in curriculum evaluation and research. Four purposes of curriculum evaluation data are identified: (1) advancement of science, (2) curriculum revision, (3) provision of data for the formulation of educational policy decisions, and (4) a method for the development and refinement of educational theory. Some of the limitations of existing methodologies that relate to these purposes are pointed out, and the use of naturalistic observational methods is seen as providing solutions to a few of these problems. The study is an effort to identify existing research and measurement problems which may contribute to the improvement of theory and practice in education. (Author/AE)

Naturalistic Observation as a Research Instrument in Curriculum Development

Harold Berlak

Metropolitan St. Louis Social Studies Center
Graduate Institute of Education
Washington University

Presented as part of a symposium "Participant Observation and Curriculum: Research and Evaluation" AERA, New York, February 1971.

My purpose in this paper is twofold. First, based largely on my experience in two curriculum development projects, to identify a number of problems and raise questions about the usefulness of conventional instruments of educational and psychological measurement in curriculum evaluation and curriculum research· second, to argue that naturalistic observational methods appear to offer an answer to at least a few of these problems.

Though I have chosen to speak of the theoretical and practical shortcoming of existing psychometric models and measures, I want to avoid casting my remarks in terms of the debate over the abstract issue of the relative merits of hard-nosed experimental approaches versus the less quantitative, less precise and less reductionist approaches to educational and evaluative research. Debate over meta-questions can be useful: however, more often than not researchers align themselves firmly to an abstract position and respond to an issue not on its merits but on the basis of the position it appears to support. This paper I hope will not be interpreted as an attack on quantitative research. Rather it is an effort to state a few of my concerns and raise some questions.

The Uses of Curriculum Research

I begin with two simple propositions. First, research is a purposeful activity. Second, the purposes of research should govern (or at least have an important influence on) the choice of method and measures. If, for example, a research project is intended to identify

the specific human-contributed factors which have led to the accumulation of excess nitrates and phosphates in the soil, the design of the study should assuredly reflect this intent. If a social researcher's purpose is to identify the causes of student unrest in order to develop recommendations for governmental policy, the purpose of the study will help determine types of evidence to be gathered and from whom. The selection of methods and procedures for "pure" research is also governed by the nature of the theoretical issues a researcher hopes to illuminate.

That purpose governs procedure seems obvious and to labor the point may seem absurd. However, I am convinced that much curriculum research, especially evaluative research, has used methods and instruments which are not related to identifiable purposes. This may be an over-statement, but I believe that curriculum developers often gather data, not with any specific intent in mind, but to placate a funding agency, a potential publisher, or because of an implicit belief they hold that behavioral research of any kind requires the collection of quantifiable data. In general, curriculum research has been based on the research models of psychology without consideration of whether such models are instrumental to specific ends. I am not here criticizing the use of all such models: rather, I raise the question of whether they are the most appropriate means of achieving all purposes of curriculum research.

I suspect the reason for the heavy dependence on psychological models is because psychology and psychometrics are among the strongest

traditions within faculties of education. (In many places psychology is almost synonomous with social science.) If psychometric and educational measurement is seen as the most appropriate approach to the study of human behavior and if classroom research is considered to be a special case of such study, then the rest follows.

In the time that I have here today I will suggest four major reasons for curriculum developers collecting and analyzing data and then go on to suggest, on the basis of my own experience, a number of measurement and other methodological problems. Many of the questions I raise are based on my experience in two curriculum projects: however, I do have in mind a number of other curriculum projects as well, mainly within the social studies. Though I am reasonably certain that many of the same questions can be raised about curriculum research in other areas, I am not making that claim here.

Four of the purposes for curriculum developers collecting, analyzing, and interpreting data are: 1. Advancing science; 2. Revising and modifying the curriculum and its rationale; 3. Providing data for educational decision-makers; 4. Developing and refining educational theory.

The Advancement of Science

Curriculum development has as its main goal providing material which will help contribute to growth of students, and research associated with such projects should help realize this practical end.

However, any study of human behavior may contribute to the development and refinement of basic behavioral science theory whether or not this is its primary purpose. There are many examples, in the physical as well as the social sciences, where data collected for practical ends contributed to the development of theoretical knowledge in a field. If a curriculum development project has on its staff persons who have some interest in basic questions, the research program provides an opportunity for deliberately collecting data which may contribute to our understanding of human behavior and society. The by-products of practical study have in the past not only led to modification of basic theory but have contributed to methodological knowledge. Curriculum research could have a bearing on such fields as motivation, political socialization, cognition, ego and cognitive development, social influence, attitude formation, etc.

In order to relate the observables in a classroom or school to theoretical considerations, we obviously need some way of describing human behavior in such settings. Stated differently, instrumentation must be capable of discriminating a wide range of human behavior and at a level of specificity which is related to theoretical requirements of existing theory. For example, if curriculum research is to have some bearing on the basic question of how the behavior of social studies teachers influences the political values and beliefs of students, we need to have instruments which can systematically record the behavior from which we can posit antecedents and consequences of teacher behavior on this process of political socialization.

The instruments which are presently available and commonly used for describing in-school and in-classroom behavior share a number of problems which appear to me to limit their usefulness for most behavioral science investigation. Systematic observation systems, whether designed for the classroom or other kinds of group behavior (e.g. Flanders, Bales, Withall, B.O. Smith, Bellack), are extremely reductionistic. Most of them were devised from a particular theoretical point of view, to test a particular set of hypotheses, or to meet a particular set of practical needs. I am not arguing that this is improper. My point is that many of these instruments deal only with a very limited range of manifest behavior. In most cases they deal only with verbal behavior in the classroom and in all cases the behavior is placed into a very limited set of categories, usually less than a dozen, and then counted.

Perhaps if we had as many reliable observational systems as there are theoretical positions our problem would be solved. But we do not, and I am not persuaded that the required investment of effort would be worth the outcome. What troubles me particularly is the insistence that some kind of systematic observational system is absolutely necessary in order to describe the behavior within school settings.

What of other instruments, for example, standardized achievement tests and other measures which are structurally similar (personality inventories, critical thinking tests)? Manning (1969) pointed out that achievement tests in general focus on products rather than processes of behavior. He also pointed out that educational measurement has

mainly developed as a technique for evaluating outcomes and rarely describes the strategies that an individual uses in reaching these outcomes. There are important distinctions among various types of educational and psychological tests and I do not want to blur these. Nevertheless, on the basis of our experience, existing psychological and educational instruments could not meet our needs because our research problems required descriptions of the processes of behavior related to outcomes. [1]

In those instances where descriptions of processes of behavior are required, and it is clear that neither existing product measures nor systematic observational instruments are adequate, I see two alternatives open. First, is the use of a computerized storage and retrieval systems such as the General Inquirer. However, I believe there remain technical cost problems to be solved before such systems can be used other than in special circumstances. The second alternative is the use of naturalistic observational techniques of the sort described by Smith . I do not see the use of naturalistic observation as a retreat from rigor; for some kinds of research it provides the most sensitive instrumentation we have. I believe the Smith and Pohland (1969) study of CAI, other studies by Smith, as well as the studies by Solomon, Seif, and Applegate begin to demonstrate its potential as an instrument for use in basic behavioral research.

---

[1] The adequacy of measures of outcomes (attitudes, values, traits, skills) is a separate question I would like to acknowledge here, but not discuss.

Revising the Curriculum

A curriculum project usually subjects its material to a field trial with the clients for whom the material is intended. Presumably, based on data collected during those trials, the materials are changed in some way. I believe it is proper that most data collection efforts by a curriculum project are directed toward this end. This collection of data for the purpose of revising a product is what Scriven has labeled "formulative evaluation."

The Harvard Social Studies Project, with which I was associated during its first curriculum development effort, collected masses of data using a wide variety of measures [2] (See Oliver and Shaver, 1966). Whatever else might be said of the usefulness of this data collection, it did not provide the information necessary for helping the developer make specific revisions in the cases, texts, and teacher materials. Such instruments certainly gave indications of gross problems, e.g., what kinds of outcomes were and were not achieved. But in order to revise materials, one needs to have descriptions of classroom processes in sufficient detail so that one is able to make good guesses about specific antecedents and consequences of using a specific component of the curriculum materials. Although the Harvard Project developed

---

[2] Instruments included: Several project constructed tests to measure thinking competence, an adaptation of the Watson-Glaser Critical Thinking Test, Iowa Test of Educational Development #5, California American History Test, Principles of American Citizenship Test, Suilford-Zimmerman Temperment Survey, a version of the Semantic Differential, Cattell High School Personality Questionnaire. In addition, two systematic observation instruments were used; an adaptation of Bales Interaction Process Analysis, and a project constructed instrument called the Analytical Category System.

a fairly complex systematic observational system, the information summarized by the system did not provide the kind of feedback needed to alter or abandon a particular story or set of teaching strategies. In retrospect I am persuaded that naturalistic observation would have provided the type of data for the type of analyses we needed. In the Washington University Curriculum Project, the basis of naturalistic descriptions of the flux of events in the classrrom, we attempted to identify specific alterations in the materials intended for the students or teachers.

Another form of data collection which is used widely by curriculum developers is survey opinionaires. As curriculum developers we want to know the perceptions of teachers and students who have been in the program. Such feedback, although it may suggest possibilities and alternatives, does not give any independent, direct evidence of what occurred in the classroom.

In our work within the Washington University Project, we have found that our revision efforts have depended almost entirely upon analysis from the descriptions of events in the classroom recorded by one or two observers. The studies by Seif, Applegate, and Solomon, are examples of thorough, systematic efforts at the collection and analysis of naturalistic data. Although their studies are not conceived primarily as formulative evaluation studies, they do exemplify the usefulness of collection of naturalistic data for curriculum revision. Smith and Pohland (1969) is perhaps a clearer example of how a study can provide the type of data and analysis for revision which could not be collected with the use of conventional educational and psychological instruments

Providing Data for Educational Policy Decisions

The curriculum used by a school is an aspect of its educational policy, and educational policy is a form of public policy. In a recent paper (Berlak, 1970), I made an effort to distinguish between "public policy" and "programmatic" outcomes in education and make some suggestions as to the appropriateness of various research and measurement techniques for both types. I will only touch briefly on a number of issues related to how a curriculum developer can provide the kind of data which is potentially useful to an educational policy maker. The adoption of a specific policy (e.g. a curriculum) requires a judgment by the decision-makers that a given set of outcomes is desirable, and a prediction that they are likely to be realized if a specific set of curriculum materials are used within a setting. The question of what is desirable requires value choices. If the curriculum developer is to provide data related to policy decisions, he must collect data bearing on the moral questions, and on the effectiveness of the curriculum which bears directly on the moral questions in two ways; first, by evaluating the effectiveness of the curriculum in terms of expected outcomes and second, by providing data on what medical researchers might call "side effects." In other words, educational policy-makers need to know what are the unanticipated and unintended effects of a curriculum which have a bearing on both moral worth and effectiveness. For example, a given program may lead to

exceptionally high gains in mathematical achievement but it may also lead to behaviors in the classroom from which one is able to infer that the students feel themselves to be pawns. A curriculum developer may have neither intended or anticipated such a consequence.

Given the complexity of behavior in the classroom, what kinds of instruments can be used to pick up such side effects? Most achievement tests, as has been pointed out, are product oriented, and many side effects cannot be picked up by "product" measures. Even if appropriate outcome measures are available, the selection of such measures presupposes that it is possible to anticipate most of the important side effects. Systematic observational instruments may be useful for some of the side effects, but as I have noted earlier, the range of school and classroom behavior to which these instruments are sensitive is limited.

Without denying that previously cited types of measures can be useful, I am persuaded that the techniques of naturalistic observation properly used are probably the most efficient and sensitive means we have available for collecting behavioral data on unintended and unanticipated consequences. For many of the complex value questions in educational policy I believe that the naturalistic observational method is the only means we have for collecting data on side effects.

Stake (1970) distinguished the wide variety of judgment

data necessary for rendering judgments as to effectiveness, and he discussed the range of instruments useful for collecting this data. In general, on the basis of my review of existing models and instruments (Berlak 1970), I have found little to justify any confidence that the field of educational evaluation as an applied social science possesses the models, strategies, or techniques for contending with the moral component of educational decisions.

There is another set of problems which is becoming increasingly important for psychometricians. As Scriven (1969) points out,many existing tests themselves are based on a set of moral presuppositions. For example, Manning (1969) notes that achievement tests are based on adversarial assumptions. Just as curriculum materials have moral presuppositions so do tests. This raises a number of interesting questions about how the presuppositions of some tests may fit the presuppositions of a curriculum. Often achievement tests are used as though such presuppositions did not exist. Because the tests are generally constructed to meet specific kinds of institutional needs, the tests are not always consistent with the uses to which they are put. Cronbach (1969 p. 36) comments that "comparisons (competition) is a theme straight out of John Stuart Mill and Charles Darwin. But evaluation of social programs and self direction by individuals calls for absolute judgments." Increasing numbers of curriculum developers and

educators are interested in non-adversarial assumptions, yet to my knowledge measures do not exist which make such presuppositions. Certainly educational policy decisions cannot await their development, and some form of research and instrumentation is necessary now.

I would like to return briefly to my earlier point that educational research has depended very heavily on the traditions of psychology and psychological measurement. There seem to be compelling reasons to begin to examine the approaches of the other sciences of human behavior, most notably those which have a tradition of systematic naturalistic observation. In our work we have learned a great deal about the use of naturalistic observation in curriculum research, but we are novices. We are only beginning to develop the ground rules appropriate to our task. As far as I know, the use of naturalistic study as a means of collecting data related to policy decisions is not common.

I would like to emphasize that in some cases I see the naturalistic studies as an intermediate step. Quantitative research, as Smith has shown, can follow the naturalistic studies. But there are instances I believe where it may not be second best but the most rigorous approach possible given the nature of the problem.

Development and Refinement of Educational Theory

For some, educational theory is merely a form of behavioral science theory. For me, educational theory includes the moral and empirical presuppositions of educational practices. For example, the curriculum requirements and organizational patterns of a liberal arts college rest upon a set of assumptions about the nature of man, society, knowledge, and the ways man best learns and can be taught. Curriculum theory I see as one form of educational theory.

How can the data collected by curriculum developers contribute to the development of educational theory? I believe that data collected during a curriculum trial can contribute to the development and clarification of alternative educational theories. For example, on the basis of our work in the schools, and our observations, I have begun to formulate some ideas as to how we may be able to reorganize elementary schools to create a more humane learning environment. Before the Washington University Curriculum Project began its work, I had a number of general concerns about the way conventional classrooms are organized, and on the basis of the study of data we have accumulated primarily for formative evaluation, I have begun to formulate these views into an educational theory which I hope will lead to alternate forms of educational practice. For example, as I have sorted through the chapters of Mr. Applegate's study of role-play and his data, I have begun to formulate specific

propositions which schools can use to organize themselves to foster student directed learning activities. I am not suggesting that my thinking at this stage is very clear but my point is that [redacted] the richness of the data collected by naturalistic observation has the capacity to make a contribution to formulation of educational theory. Our data has reopened some basic questions about the functions of schooling within our society. Though I only raise this issue briefly, I do not want to underestimate its importance. Professional educators including researchers are often so enmeshed in existing practice that they generally have not explored alternatives. I am convinced that the naturalistic observation describing school as it is as Goodlad (1964) suggested can contribute to the search for alternatives in which we do not simply repeat all the mistakes of the past.

Summary

I have pointed out four major purposes for curriculum research and attempted within each to raise a number of questions, to point out some of the limitations of existing methodologies, and to suggest a number of ways that naturalistic observation can contribute to the solution of these problems. I do not intend this paper be a defense for a particular methodology. Rather, it is a modest effort to identify a few research and measurement problems which may contribute to the improvement of theory and practice in education.

Bibliography

Berlak, H. Values, Goals, Public Policy and Educational Evaluation. Review of Educational Research. Vol. 40, No. 2. 1970.

Cronbach, L.J. Validation of Educational Measures. In Proceedings of the 1969 Invitational Conference on Testing Problems: Toward a Theory of Achievement Measurement. Princeton N.J. Educational Testing Service. 1969.

Goodlad, J.I. Curriculum: State of the Field. Review of Educational Research. Vol. 39, No. 3. 1969.

Manning, W.H. The Functions and Uses of Educational Measurement. In Proceedings of the 1969 Invitational Conference on Testing Problems: Toward a Theory of Achievement Measurement. Princeton N.J. Educational Testing Service. 1969.

Oliver, D.W. and Shaver, J.P. Teaching Public Issues in the High School. Houghton Mifflin Co., Boston, 1966.

Smith, L.M. and Pohland, P. Education, Technology, and the Rural Highlands. In Howard Russell (ed.) Evaluation of Computer Asserted Instruction Program. St. Ann, Mo. Central Midwestern Regional Educational Laboratory, Inc. 1969

Stake, R.E. Objectives, Priorities and Other Judgment Data. Review of Educational Research. Vol. 40, No. 2. 1970.

Scriven, M. Discussion. In Proceedings of the 1969 Invitational Conference on Testing Problems: Toward a Theory of Achievement Measurement, Princeton N.J. Educational Testing Service 1969.

Curriculum materials for grades 4-6, developed by the Washington University Elementary Social Studies Project of the Metropolitan St. Louis Social Studies Center, are being published by Singer Division of Random House, New York.