DOCUMENT RESUME

ED 124 584 TM 005 340

AUTHOR Gephart, William J.
TITLE Problems in Measuring the Degree of Implementation of an Innovation.
PUB DATE [Apr 76]
NOTE 12p.; Paper presented at the Annual Meeting of the American Educational Research Association (60th, San Francisco, California, April 19-23, 1976)

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS *Educational Innovation; *Evaluation Methods; Measurement; Problems; *Program Development
IDENTIFIERS *Implementation

ABSTRACT

This paper argues that measuring the degree of implementation is important if causative statements about effects of an innovation are to be made. It identifies four classes of problems which impede such measurement: the purpose problem, the local adaptability probelm, the scalor problem, and the innovation completeness problem. An understanding of these four classes of problems should lead to better efforts at measuring the degree of implementation of an innovation by making one conscious of: (1) the need to specify the purpose for measuring and to focus work on that purpose; (2) the probability and acceptability of local adaptions and the need to assess their appropriateness given the local circumstances; (3) the futility of a single measure and the advantages of a profile in measuring degree of implementation; and (4) that responsibility for outcomes of the use of an innovation rests both with the developer and the implementor. (Author/DEP)

US DEPARTMENT OF HEALTH
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

PROBLEMS IN MEASURING THE DEGREE OF IMPLEMENTATION OF AN INNOVATION*

by

William J. Gephart
Director of Research Services
Phi Delta Kappa

The rationale for the importance of measuring the degree of implementation of an innovation has been made at least implicitly through comments in this session and through writings about the research and evaluation process. That rationale is the basis for "process evaluation" in the CIPP evaluation model (Stufflebeam, et al. 1971) and for "implementation evaluation" in the evaluation model described by the UCLA Center for the Study of Evaluation (Klein, et al. 1971). Briefly stated, that rationale says that if you don't know what happened to a group on which you have outcome measures, you cannot explain what caused the observed outcomes. A pretest and a posttest by themselves are insufficient. We know full well that frequently the experiences we want to occur between them are not carried out as planned. In fact, in some instances I've heard of, those experiences were not even initiated.

A change in the intended experiences (read treatments) is often fatal to a research effort for such a change makes the independent variable tested different from the one that the review of literature indicates needs to be tested. It is just as fatal to an evaluation. Here a failure of the staff to carry out the intended experiences (read sometimes as an innovation) that goes undetected by the evaluator results in mis-information delivered to the decision maker. For example, if a set of classes is involved in an evaluation of the experience chart approach to teaching reading and the teachers

---

*A presentation at the annual meeting of the American Educational Research Association, San Francisco, April 22, 1976

involved seldom elicit stories from the children, the pre- and posttest differences in means will not help the decision maker know how effective that innovation is in his or her setting. Measuring the degree of implementation of an innovation is required if we hope to say anything about its effects or its worth.

It is much easier to say we must measure the degree of implementation than it is to do it. At least four classes of problems abound in such efforts. Those four are the purpose problem, multiple scale problem, a local adaptibility problem, and an innovation completeness problem. The remainder of this presentation will attempt to describe the nature of those problems and the complications they create.

The Purpose Problem

What is our purpose? What accomplishment will we contribute to by measuring the degree of implementation of an innovation? Ralph Tyler (1969) has listed and described a half dozen general meansurement purposes and strongly implies that the measurement procedures will vary depending on the purpose being served. The maxim, form follows function, applies here. The same measurement form will not accomplish all functions. If we want to diagnose implementation difficulties we would use a different measurement procedure than if our function is to provide information useful in an adopt/adapt/-reject decision. The corollary of that maxim also applies, function follows form. That is, if we use the same measurement procedures for all purposes, we will not efficiently and effectively serve all the purposes. Rather we will serve only one purpose.

The importance of function or purpose is recognized in Nadler's work (Nadler & Gephart, 1972) on the design or development process. Until the purpose or function to be served is specified, it is a mistake to design or select tools or procedures for doing the work. Nadler operates on the maxim, function is first.

Guba (personal conversation) gives some very practical help related to finding and stating our purpose that is applicable in measuring the degree of implementation of an innovation. He suggests that you put yourself into the future and pretend that the measurement has been completed and that it did the job needed. Now, Guba suggests, answer the question, "What has been accomplished as a result of measuring the degree of implementation of the innovation?" The answer to that question usually is your purpose.

Before marching off to measure and serve that purpose, Nadler (Nadler & Gephart, 1972) implores us to check first to see that the purpose needs to be served. He urges the development of a purpose or function hierarchy. This is done by starting with the purpose identified through Guba's question; pretending that function has been accomplished; asking what higher purpose would be our concern; and recycling this questioning until the hierarchy is extended as far as possible. Nadler says that the appropriate function or purpose on which to focus is the one that is least restrictive or limiting to the system.

The purpose problem is an impediment to measuring the degree of implementation of an innovation. From the discussion above it would seem it manifest itself in the form of: (1) failure to delineate the purpose or function to be served by the measurement effort; and (2) failure to use measuring procedures appropriate to that purpose.

The Local Adaptability Problem

The second problem in measuring the degree of implementation seems to this observer to have resulted from a myth now embraced by educators and the American public. That is the myth of the teacher-proof system. Lots of people have contributed to this myth. Some of us have wished for the teaching machine that would do the perfect instruction job. Others of us have tried to develop them. Still others (noteably some of our bureaucratic leaders in Washington) have demanded that we "validate the transportability" of educational products we create.

This myth of universal applicability (another way of saying validated as transportable) is folly, utter folly! Nothing that has been created in the history of American education has been shown to be universally applicable. No text, no teaching machine, no test, no teaching procedure has accomplished this feat. We have enough of a burden trying to create tools and procedures that will work in the variability found in one setting, let alone requiring that our creations be capable of meeting the situational variables in all settings!

This objection to the myth of universal applicability is not simply the frustrated moan of an unsuccessful developer. A specific product or procedure is developed for a particular purpose or function. And typically, purposes or functions differ from setting to setting. Nadler (Nadler & Gephart, 1972) reports being asked to assist a hospital staff develop a medical records library system. They created one that greatly satisfied the hospital staff. Three months later Nadler had an identical request from another hospital. Most people would use the system developed three months earlier. Nadler did not. Rather, he employed the design process as if the first medical records library work had not been done. In working with the second hospital, it was quickly apparent that the function or purpose for the medical record library system was not the same from one hospital to the other. Thus, a somewhat different medical record library system was designed for the second hospital.

Some areas of the business and manufacturing world recognize that for a product (or procedure) to be effective it must be locally adaptable. For example, when you purchase an automobile you exercise a number of options to fit the vehicle to your purposes and desires. Still other "local adaptations" are involved as indicated in a charge you pay called "Dealer set-up charges." This includes, for example, adjusting the fuel-air-spark parameters so that an automobile assembled in Detroit (500 feet above sea level) can operate efficiently in Denver (5,000 feet above sea level). The same adaptability can be seen built into other products, desks come with some adjustment to the length of the legs, so that accomodation to uneven floors is possible. Desk chairs come with a height adjustment, etc.

Innovations that have any complexity are systems with numerous component parts. The ideal system would be one which has the needed number and type of components universally required <u>and</u> the needed number and type of component parts that would permit the local adaptation required to fit the difference in purpose found in the settings in which it would be used. Achieving that perfectly adaptable product or procedure is unlikely, however. First, we seldom know enough in a design effort to create <u>all</u> the needed component parts. (System analysis people refer to this as the degree of system wholeness.) As a result, we "patch-around" unavailable components. Second, the tool or procedure to be created, if it has any complexity, has a set of required knowledge and skills for its effective operation. Thus, the knowledge and skills possessed or easily developed by the personnel who will use the innovation become an upper limit for a use in a specific setting. Because of these two factors, any given product or procedure is modified as it is used.

Given the need for and fact of local modification of an innovation, the measurement of degree of implementation is complicated. The problem becomes one of defining anticipated, actual, and appropriate adaptation. Preset and rigidly structured measuring techniques cannot be used effectively in situations in which flexibility and modifiability are the rule.

When evaluation generalizability is sought (and measuring the degree of implementation is some instances is process evaluation), our efforts should be patterned on the consumer products model as illustrated in the continuing work of the Consumers' Union. (Consumers Report; Gephart & Potter, 1976) Consumers Report shows attention to two types of decision appropriate information. The first type consists of those questions for which there is no correct answer (for example, do you want to buy a car in which you will "feel the road" or in which you will float?), questions settled by personal values or situational conditions. The generalizable evaluation report (and thus, the measurement of the degree of implementation) should alert those interested in the innovation to the set of questions on which they have options to exercise. The second type of information consists of those items that are constants, that are not situationally variable. Both types of information are necessary to communicate about an innovation's worth to potential users.

The Multiple Scale Problem

As indicated earlier, any innovation that has some complexity is a system with numerous component parts. Assessing the degree of implementation requires observations or measurements on a number, if not all, of these components. Some of these components will be observable in categories yielding nominal data (for example, were all the necessary kinds of equipment -- desks, chairs, books, etc. -- assembled before the innovation?). Still other components of the innovation will be of such a nature that they may be measured in ratio scales (for example, how much time was devoted to component X of the innovation?). Ordinal and interval measures are also likely to be involved if the innovation is complex.

The use of different measurement scales creates a difficulty in summarizing data. It is conceptually impossible to combine nominal, ordinal, interval, and ratio data without losing some information. Thus it is impossible to get a single score that clearly describes the degree of implementation of an innovation if different scales of measurement are involved. It is relatively easy to get a score or scores on the various components and thus to present a profile of the degree of implementation. But combining those profile items to a single summary descriptor of the degree of implementation requires value judgment about the relative value or weighting for the different items in the profile. And, as we all know, value judgments vary from person to person thus creating differences in perception of the overall quality of a given profile.

Given the statements made earlier related to the need for adaptability in educational products and procedures, a profile would seem to be a more logical and beneficial way of communicating about implementation of an innovation than a single score.

The Innovation Completeness Problem

Frequently a new product or procedure in education appears complete to its developer but proves to be incomplete in another setting. This incompleteness is the fourth category of problems in measuring the degree of implementation of an innovation. Incompleteness of one sort has already been alluded to in this presentation. That is incompleteness that occurs due to our lack of knowledge or ability to create some of the components of the ideal product or procedure. As a result, developers "patch around" the missing components to create a feasible product or procedure. Developers cannot be faulted for this type of incompleteness. We cannot ask that development efforts be suspended until all the necessary knowledge or ability to create is in. It does stand as a developer fault, however, if potential users of the innovation are not alerted in advance regarding the points of incompleteness.

A second form or source of incompleteness is more subtle and invidious. That is the indispensible person problem. Developers have created products and procedures which work when they are involved but not when it is turned over to someone else. In such cases the crucial element in the innovation seems to be the style of operation of the indispensible person. For example, several years ago a group was told that an individual had created the perfect way of teaching reading. After some discussion it was learned that the predicted success had been demonstrated when, and only when, the developer was the teacher. To the credit of the developer, his role as a part of the innovation was recognized. Others are not that observant.

One of the questions that should be asked about an innovation then is, "What are its points of incompleteness?" This is perhaps best expressed in the language of the systems analysts. Such people consider a tool or procedure as a system and the components of it as subsystems. In general systems theory, it is readily accepted that all of the subsystems interface (interlock or connect) in a manner which maximally serve the overall system's function. Systems analysts speak of the concept of wholeness in this respect.

Two techniques which help pinpoint innovation incompleteness are flow-charting and PERT diagraming. These are aides both to the practioner and the person charged with measuring the degree of implementation. Flow-charting involves the development of a chart that sequences and interrelates actions of various sorts and decision points. A variety of geometric figures are involved and their different shapes present relevant meaning. (For example, a rectangle generally indicates some kind of activity; diamonds represent decisions and specify a set of alternatives; circles are connectors, etc. Templates for these symbols are available at most drafting suppliers.[9] PERT is the acronym for Program Evaluation and Review Technique. PERT is an analysis and review procedure created as a management tool for the Office of Naval Research at the time the Polaris Missile System was being created. Cook (1965) as described PERT and its applicability to education. Central to PERT is the creation of what is called a PERT chart in which arrows represent activity and circles represent events. The PERT chart shows the sequence and interrelation or interdependencies of events and subsequent activities. If either a flowchart or a PERT chart is made for an innovation, the points of incompleteness are more liable to be observed than if a purely verbal description is presented.

Incompleteness in an innovation hampers measurement of the degree of implementation <u>if it is undetected</u>. In those cases the measurer <u>assumes</u> that event A will follow activity A which will be followed by activity B and event B, etc. A point of incompleteness invalidates the assumption. By assuming that the innovation is complete when it isn't, we shift the responsibility for discrepancy between what the innovation is claimed to accomplish and what it does accomplish from the developer to the implementer, a shift that is not logically warrented.

Summary

This paper has argued that measuring the degree of implementation is important if we want to make causative statements about effects of an innovation. It has identified four classes of problems which impede such measurement: the purpose problem, the local adaptability problem, the scalor problem, and the innovation completeness problem. An understanding of these four classes of problems should lead to better efforts at measuring the degree of implementation of an innovation by making us conscious of: (1) the need to specify our purpose for measuring and to focus our work on that purpose; (2) the probability and acceptability of local adaptions and the need to assess their appropriateness given the local circumstances; (3) the futility of a single measure and the anvantages of a profile in measuring degree of implementation; and (4) that responsibility for outcomes of the use of an innovation rests both with the developer and the implementor.

References

Cook, D. L., Program Evaluation and Review Technique -- Applications in Education. Cooperative Research Program Monograph 17: Washington, D.C. 1968 (ERIC Document No. ED 015 533).

Gephart, W. J. & J. W. Potter, "The Generalizability Problem in Evaluation and a Borrowed Solution." Unpublished paper, Phi Delta Kappa: Bloomington, Indiana, 1976.

Klein, S. P., et al., Evaluation Workshop I: An Orientation., Center for the Study of Evaluation, UCLA, CTB McGraw-Hill: Monterey Park, California, 1971.

Nadler, G., & W. J. Gephart, "The Process of Development." Phi Delta Kappa Occasional Paper #15: Bloomington, Indiana, 1972.

Stufflebeam, D. L., et al., Educational Evaluation and Decision Making. F. E. Peacock, Publisher: Itasca, Illinois, 1971.

Tyler, R. N. "The Purposes of Assessment." In W. H. Beatty (ed.) Improving Educational Assessment and An Inventory of Measures of Affective Behavior. Association for Supervision and Curriculum Development, NEA: Washington, D.C., 1969.

of complexity, a theory should reflect this knowledge. Research into the factorial complexity of APT forms would also contribute to theory development.

8. There is a need for creative development of new forms of APT that may alleviate some of the measurement shortcomings that have been discussed. Educational measurement specialists funded to explore such creative alternatives would contribute new knowledge that would have immediate use for public school testing.

Applied Performance Testing has great appeal for measuring task performance in the public schools. There is much work to be done to refine the concept and improve on our techniques. I believe the effort is worthwhile and expect to see comparatively great advances in APT in the near future.

REFERENCES

Cronbach, Lee J. Essentials of Psychological Testing. New York: Harper and Row, 1970.

Fitzpatrick, Robert and Morrison, E. J. Performance and Product Evaluation. In R. L. Thorndike (Ed.) Educational Measurement (2nd Edition). Washington, D.C.: American Council on Education, 1971.

Gagne, Robert M. Simulations. In Robert Glaser (Ed.) Training Research and Education. Pittsburgh: University of Pittsburgh Press, 1962.

Glaser, Robert (Ed.) Training Research and Education. Pittsburgh: University of Pittsburgh Press, 1962.

Harris, Chester W., Alkin, Marvin C. and Popham, W. James. Problems in Criterion-Referenced Measurement. Center for the Study of Evaluation Monograph Series in Evaluation, No. 3. Los Angeles: UCLA Center for the Study of Evaluation, 1974.

Osborn, William A. Developing Performance Tests for Training Evaluation. In James R. Sanders and Thomas P. Sachse (Eds.) Problems and Potentials of Applied Performance Testing. Portland, Oregon: Northwest Regional Educational Laboratory, 1975.

Panitz, Adolf and Olivo, C. Thomas. The State of the Art of Occupational Competency Testing. New Brunswick, N.J.: Department of Vocational-Technical Education, Rutgers University, 1970.

Sachse, Thomas P. and Sanders, James R. A Look at Applied Performance Testing in Education. Portland, Oregon: Northwest Regional Educational Laboratory, 1975.

Smith, N. L. The Study of Criteria in Educational Evaluation. Unpublished Doctoral Dissertation. Champaign: University of Illinois, 1975.

Webb, E., Campbell, D. T., Schwartz, R. D., and Sechrest, L. Unobtrusive Measures: Nonreactive Research in the Social Sciences. Chicago: Rand McNally, 1966.