DOCUMENT RESUME

ED 312 315 TM 014 122

AUTHOR Eash, Maurice J.; And C...
TITLE Determining Outcomes for Ev luation of a National Science Foundation Workshop.
PUB DATE Oct 89
NOTE 15p.
PUB TYPE Reports - Evaluative/Feasibility (142) -- Reports - Research/Technical (143)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Biology; Chemistry; Evaluation Methods; *Evaluation Problems; High Schools; *Program Evaluation; Science Education; *Science Teachers; *Secondary School Teachers; Teacher Improvement; *Teacher Workshops
IDENTIFIERS *National Science Foundation; Our Class and Its Work Questionnaire; *Outcome Oriented Evaluation

ABSTRACT

Problems in conducting an outcome evaluation under uncertain knowledge of effects are addressed. This evaluation was designed to obtain outcome assessments of a National Science Foundation Workshop, which was working with 20 high school chemistry teachers and 20 high school biology teachers to stimulate effective teaching, assist teachers in developing instructional modules, and provide a broader network of scientists to relate to high school science activities. The initial eval ation design called for self-reports of program impact from participants, but hard data, in terms of test scores, became necessary to document impact more directly. Three measures were selected to gather data on workshop impact: (1) a questionnaire determining the reactions of participants; (2) a questionnaire examining the utility for teachers of activities in the workshop as translated into their teaching; and (3) an instrument, Our Class and Its Work (OCIW), developed by M. J. Eash and H. Waxman (1983), which was used to gather information from students of workshop teachers (N=442) and students of control teachers (N=384). Students' perceptions of classroom approach paralleled teacher self-reports; when teachers reported a change in classroom behavior, OCIW student data suppored a change. The problems in outcome evaluation illustrated by this study indicate the need to rely on more than just test scores in evaluating program impact. (SLD)

Determining Outcomes for Evaluation of A National Science Foundation Workshop

Maurice J. Eash
William Hagar
Walter Weibrecht
University of Massachusetts/Boston
Boston, MA 02125

October, 1989

Determining Outcomes for Evaluation of A National Science Foundation Workshop

A number of years ago in a speech before graduate students Paul Lazarsfeld, the Sociologist, observed that all evaluation studies had four components: What outcomes was the study interested in? How were these outcomes measured? What comparisons were made? What specification and generalization of the outcomes could be drawn? All evaluation in Lazarsfeld's paradigm of evaluation studies point toward given outcomes (effects) and I suspect that this is as true for most evaluations today as it was over thirty years ago when I first encountered his observation. The assumption behind such exclusively oriented outcome evaluation is that project director and evaluator can know in advance and identify with certainty, if not certitude, the effects of treatment as manifest in outcomes. The case that I shall examine in this paper is that of conducting an outcome evaluation under uncertain knowledge of effects. The evaluation was designed to obtain outcome assessments of an NSF workshop for high school science teachers, entitled "Applications of Basic Science in Industry and Society to Enhance Secondary School Science".

This workshop-seminar was part of the Teacher Enhancement effort funded by the Directorate in Science and Engineering under the National Science Foundation. It was designed to work with 20 Chemistry and 20 Biology teachers in a three week summer workshop and a year long series of six seminars. The objectives as submitted with the project were:

"1. To stimulate effective teaching approaches to theories and concepts in Biology and Chemistry for secondary science teachers who are teaching in several science disciplinary fields.

2. To build curriculum units from industrial and societal applications of conceptual and syntactical science using teaching strategies geared to the range of students abilities.

3. While the major purpose of the project is teacher enhancement a secondary focus is to provide training for teacher-leaders who will be identified and encouraged to give

leadership in their schools and to science education in the Commonwealth.

4. To provide lesser prepared Biology and Chemistry teachers (especially those teaching in other than their major) an opportunity to build enriched curriculum units and develop more effective teaching strategies.

5. To develop long-range networking opportunities for secondary science teachers in the metropolitan area to work together and profit from a continuous exchange of new materials and teaching strategies.

6. To provide impetus and incentive for development of long term collaborative relationships between secondary science teachers, college and university scientists, and industry and government affiliated scientists."

These six objectives had three central foci: to stimulate more effective teaching, to assist teachers with developing instructional modules and to provide a broader network of scientists to relate to high school science activities. To achieve these objectives a series of activities were designed which included visits to 10 industries, agencies and natural sites; the development by teams of teachers of instructional modules; laboratory work and lectures; all of which culminated in units of instructions which were taught in the academic year following the summer workshop. In the six academic year seminars teachers' teams presented their efforts, described the implementation in the classroom, and received critiques on their work. The completed modules were written up, printed and distributed to all participants.

An evaluation design which had been sketched in the original project called for an external evaluator to gather data from participants on their reaction to the workshop, i.e., did it accomplish the six objectives, and to verify these responses through interviews with teachers and administrators in their schools. After the first summer workshop was finished, our program officer at NSF said the Foundation for Congressional support needed to receive data that more fully documented the "impact" of the workshops on the participants. When queried on what was considered impact data - he said, "test scores". This set off a scramble to redesign the evaluation to be more specific and responsive to NSF's understanding of outcomes. It was made quite clear that self report data, on participants' enthusiastic response to the workshops, was no longer acceptable as evaluation outcome data. Hard evidence, "test scores", was required.

In resolving this evaluation design problem, we encountered a range of philosophical and technical problems which are posed by outcome evaluation. These are presented under four questions:

1. Can the outcomes of the impact of the workshop be identified?

Given the mandate to produce "test scores", and specifically achievement test evidence that would demonstrate that the teachers had learned "more science", a quick search was made of standardized tests that were available. As the teachers were preparing their own modules from their summer activities these deviated from the standard textbook curriculum. It took limited . detective work to determine that the achievement tests available for use with either students or teachers lacked curriculum validity. We also came to the conclusion that a logical linear model of evaluation. Objectives specified --> activities performed --> outcomes relating to objectives measured --> conclusions drawn from the analysis of these data was too reductionistic of the dynamics involved in this experience - centered approach to teacher enhancement and curriculum improvement. Moreover, the workshops objectives as drawn in the original proposal were broad and general - unprescriptive as to curriculum treatment. Indeed, this audience scarcely needs to be reminded that achievement tests have severe limitations on measuring goals of curriculum that encompass more than direct knowledge of subject matter - interest, curiosity, problem solving skill, critical analysis - to name a few domains that are largely missing from the psychometric tasks. We early-on determined that we must cast our evaluation net much broader to detect the impact of the workshop-seminar treatment.

As the result of this limitation of extant tests, it was decided to seek a different measures that were both more sensitive and representative of the intents of the workshop. In this decision three types of measures were selected to gather data on the workshop treatment:

(1) A questionnaire prepared by the external evaluator on participants' reaction to the different components of the workshop and to document the implementation of an instructional treatment.

(2) A questionnaire prepared by the internal evaluator and administered at the conclusion of the workshops and seminars asking teacher participants about the utility of the activities in the workshop as translated into their teaching, e.g. had they used any of the sites visited in their own classroom field trips.

(3) A third instrument the <u>Our Class and Its Work</u> (OCIW) was administered to all participants and to two separate control groups.

This instrument has the advantage of gathering student perceptions of teacher's classroom behavior that teaching research has shown to effect student achievement.(Waxman & Eash, 1983) The <u>OCIW</u> (Eash & Waxman, 1983), a learning environment measure, is composed of eight scales of forty items which describe teaching behaviors . A ninth scale of ten items was added to gather data on students' reactions and attitudes toward science. Students respond to items on a four point scale of strongly agree to strongly disagree. The scales and samples items are presented in Table I.

TABLE I
SCALES AND DESCRIPTION OF OCIW
[Reliabilities in ( )s reported in Chronback Alphas]

<u>Didactic Instruction(.87)</u>
Implies that the teacher controls and directs the instruction for all students in the class.
Sample item:"Our teacher lets us do things our own way."
<u>Enthusiasm (.92)</u>
Considers the extent to which a student sees the teacher exhibit excitement and interest in teaching.
Sample item: "We try new and different things in our classroom."
<u>Feedback (.91)</u>
Describes the extent to which the teacher respond to students answers and provide students with feedback about their schoolwork.
Sample item:"Our teacher carefully checks all our work."
<u>Instructional Time (.86)</u>
Refers to the time students spend in learning.
Sample item:"We are always working in our class."

Opportunity to Learn (.92)
Indicates how well the teacher provides opportunities for all students to learn or cover criterion material.
Sample item:"Many students do not finish their work."
Pacing (.92)
Deals with whether or not the classroom work is at the appropriate level of difficulty for students in the class.
Sample item:
"Our teacher spends too much time going over work."
Structuring Comments (.91)
Refers to whether the teacher provides overviews at the beginning and end of instructional sequences and whether students understand.
Sample item:"Our teacher often reviews yesterdays work."
Task Orientation (.84)
Indicates the extent to which the classroom is businesslike.
Sample item:"We always have an assignment to work on."
Attitude Toward Science (.77)
Indicates students attitudes toward importance of science in society, the scientific method in their lives, and science as a chosen career.
Sample item:"Scientists improve our lives."

In the absence of specific objectives detailing subject matter to be learned, and in the absence of achievement tests which would measure with specificity and accuracy the teachers' achievement of science taught in the workshop, a more indirect measure of students' perception was used as the principal outcome measure. However this measure had been validated on criteria linking teacher behavior to increased or decreased student achievement. This indirect measure of impact was related to achievement outcomes and did not depart from the original more open-end objectives of the project. The results of comparisons of the NSF Workshop and comparison groups are carried in Table II.

TABLE II

COMPARISONS OF IMPACT OF WORKSHOPS AND SEMINARS ON 1988 PARTICIPANTS' CLASSROOM TEACHING BEHAVIORS AND STUDENT ACHIEVEMENT (OCIW)

GROUP COMPARISONS	N	t	Level of Significance
1. PRE and POST TEST:			
NSF Participants' Classes			
Biology	136	17.14	.01
Chemistry	306	6.93	.01
Combined Biology & Chemistry	442	12.34	.01
2. 1988 NSF PARTICIPANTS' CLASSES COMPARED TO NON NSF CLASSES:			
Combined Biology & Chemistry 1988	442	21.56	.01
Combined Non NSF classes	384		
3. 1988 NSF PARTICIPANTS (Post Test May 1989) COMPARED TO 1989 NSF PARTICIPANTS (Pre Test May 1989):			
Combined Biology & Chemistry 1988	442	19.31	.01
Combined Biology & Chemistry 1989	630		
4. EXPERIENCED TEACHERS 1988 COMPARED TO INEXPERIENCED TEACHERS	442	14.85	.01
5. BOYS COMPARED TO GIRLS, 1988 VS WORKSHOP CLASSES	442	2.12	NS

2. Is the outcome the specific results of the treatment and if so is it specific to components of the treatment?

The traditional issues of assessing outcome effects in an evaluation design are embedded in this outcome problem.(Lipsey, 1988) To control for the effects of other variables, three types of <u>OCIW</u> data on experimental and control groups were gathered:

(1) pre and post measures were gathered on the first year participants' classes;

(2) data on a similar science class of a non-participating teacher in the same school were gathered;

(3) data on a science class of a non-participating teacher at the end of the year were gathered.

Quite large sample sizes were obtained in each comparison (Ranging from 136 to 630). Due to the large sample sizes on the controls and experimental populations, the workshop treatment would become the major measurable difference between the two groups, thus lending credence to the belief that these differences were the results of treatment and not variables extraneous to the study.

Components of the treatment and their success in achieving specific measurable outcomes were more difficult to ascribe. In the measure on utility of workshop experiences in classroom enhancement, teachers were asked which workshop experiences they were able to incorporate directly into their classrooms - as one measure of outcomes being a specific result of treatment. Of the fourteen activities in the workshop treatment that were checked for direct classroom outcomes, nine were cross validated in the <u>OCIW</u>. (see Table III)

TABLE III

UTILITY OF WORKSHOP EXPERIENCES ACTIVITIES IN CLASSROOM ENHANCEMENT (Participants Self-Report)

Activity or Experience	Percentage
Have taken a field trip(s) that replicated those in NSF Workshop	42
Have used new equipment in my teaching this year.	64
Have used a different method to teach a concept or major learning *	75
Have introduced new materials into regular curriculum *	93
Have had students plan and organize cooperative group projects *	43
Have changed my approach to incorporate more student planning in new curriculum	21
Have brought more resource people from industry into my classroom	29
Have shared my project experience and material with other teachers	93
Have included in classroom work direct applications of science concepts in industry *	82
Have observed increased student interest in science as a career this year *	54
Have increased use of questioning in my class this year *	57
Have found students with a wider range of academic abilities more interested in science this year *	50
Have focused specific lessons on the importance of science in society and specific community *	75
Have expanded the science content taught into new areas this year *	82

* Item verified by student observation data from OCIW

The activities were not specifically delineated in the original objectives and were an outgrowth of treatment as the workshop progressed. This reflected a problem in outcome evaluation where outcomes are not always known in advance, an issue on which more will be said later.

3. Can the outcomes be assessed directly?

Our linear models of evaluation and the concern for direct unequivocal evidence leads users of evaluation to long for direct measures, i.e., in student achievement primarily test evidence. We could not measure with disciplinary tests the direct effects of the workshop's treatment so we settled for use of a surrogate variable, teacher behavior. However, we did use teacher behavior in the context of promotion of student learning. That is, we used an instrument that had been validated as detecting those teacher behaviors that increased student achievement.

The selections of surrogate variables is important but it relates to a more significant evaluation consideration, the evolution of objectives as the project progresses. In short, we get smarter as we gain experience in a project. For example, we over-estimated the effect that teacher prepared materials would have on reducing the amount of didactic instruction and under-estimated it on producing gains in the other 7 scales. In addition, the project directors did not anticipate that the treatment would have much the same effect on boys and girls and reduce the gender differences that have generally prevailed in other studies of gender differences in science achievement and attitudes toward science.

What gives us most confidence in our indirect measure of outcomes, through student perceptions, is the consistency of the three sources of data. Student perceptions paralleled teacher self report data whether it was descriptive or an opinion. When teachers reported the changing of classroom approach the OCIW student data substantiated this response.

4. Can the outcome results be specified (variable relationships) and generalized to other groups?

This outcome question brings us back to an earlier concern on linking the cause and effect on a linear model. Our treatment has been documented in videos and the text activities can be replicated by others. In these respects it is both specific and generalizable. As defined in the original objectives it is not. Our learnings of what was effective came as the workshop faculty interacted with participants and received feedback from the context of conducting the planned activities. For example there was a need to change some of the sites. In addition in the second summer workshops the site visits were restructured and became more focused and in-depth than in the first summer.

The evolutionary nature of instructional treatments that are broadly based means that evaluative measures for outcome purposes if too narrowly conceived, especially where objectives evolve as the project proceeds, will fail to detect important outcomes. Thus the call for achievement gain scores based on standardized tests may very well provide outcomes which report impoverished results when what is involved is impoverished measurement.

This problem of outcome evaluation was brought home very sharply to me a number of years ago when I was called in to evaluate a Title III, teacher development project. While doing my data gathering I was asked specifically to look in on a project that focused on support for beginning teachers in several high schools. The project released a very strong creative Biology teacher who worked with a group of 13 new teachers. An outgrowth of a concern for the high attrition of new teachers in this district of large suburban high schools, the project had been underway for 6 months when I came on the scene. As in many cases when an evaluator is asked to do an add-on once into the assessment, this one had a special interest feature to it. The Title III project had an internal evaluator who was charged with gathering data on the effectiveness of the 15 projects, ostensibly to build a plausible case for renewal of funding. What I did not know when asked to look closely at the new teacher support project was that the internal evaluator and project director had come to fisticuffs over evaluation and the internal evaluator had been closed out of the scene. When I met with the project director, who proved to be an engaging chap with much teaching moxie, he gave me a quick explanation of the conflict. "He (the evaluator) kept asking me to tell him what outcomes I wanted - stated as objectives - so he could develop an evaluation design." When I stated that my outcome was to retain new teachers in the system, he said. "that's too vague an objective for evaluation and I had to get more specific." The project director responded that he was busy with planning his program and having daily meetings with the teachers. The internal evaluator said I had to have specific objectives for my activities so he could evaluate them, or he would have my funding cut-off. Already tense from venturing into new ground, this was too much for the project director-teacher to take and he proceeded to reenact an academic version of the golden gloves with the internal evaluator.

Our difficulties in assessing the outcomes of this NSF workshop were addressed by Jonathan Z. Shapiro in his writing. Early on in his career he pointed out the need to attend to both methodological and social concerns in an evaluation.(Shapiro 1984) An evaluation Shapiro claimed was unique not in purpose but in environment.(Shapiro 1985) Hence the evaluator does not generate achievement "test scores" as automatic outcome measures. The unquestioning use of such standardized measures has certainly not raised the confidence level of decision makers in the sophistication of evaluators nor added to the utility of their findings.

Respecting the dynamic nature of the environment influencing the treatment as it did, negated the belief that an evaluator could draw up a rigorous design sufficiently explanatory directly from the original objectives of a proposal. Neither project designer or evaluator were this farsighted at the origin of the proposal. If one could accept the ambiguity of incompleteness and engage without reservation in the search for partial knowledge, then evaluator, decision makers, and project subjects could be employed in the evaluation design. From the conception of dynamic environmental dependence of evaluation and the evolving of objectives as a project progressed Shapiro became committed to participant evaluation - where participants became actively involved in the evaluation particularly seeking unanticipated outcomes in the data gathering.(Shapiro 1987) In addition he was able to demonstrate how involvement could be used without sacrificing methodology rigor and in fact would strengthen the evaluation and enhance the breadth and use of the finding.(Shapiro 1984a)

What was central to any evaluation Shapiro, held is a commitment to a concept of social justice. Thus in conducting an evaluation one is required to search for issues of larger meaning than simply the easy to measure - forestalling the rush to the standardized achievement test. (Shapiro 1986) In this belief our evaluation of the NSF Workshop gives attention to students thoughts and opinions about science and society, science and citizenship, and science and a constructive career. As our teachers read these findings and see their effect, "impact", on student outcomes they will understand more thoroughly than ever the moral source of teaching. By moral I mean, the shaping and directing of individuals lives which, without the unique experience in science, might never have ventured into these domains of knowledge and thought, or more extremely, chosen to commit one's life-work to science.

Shapiro in his life and work saw outcome evaluation in the larger sphere of human endeavor. Evaluation of outcomes should be conducted to assist projects to contribute to broader choice (essentially to greater freedom), and, therefore, a more just society, and a more humane spirit. That is indeed a fitting legacy for any professional career and one all evaluators could aspire to.

Footnotes

Waxman, H. and Eash, M. (1983) "Utilizing Students' Perceptions and Context Variables to Analyze Effect Teaching: A Process-Product Investigation. " Journal of Educational Research 76(6) pp.321-4.

Eash, M. and Waxman, H. (1983) Our Class and Its Work User Manual pp.1-12.

Lipsey, M. (1988) "Practice and Malpractice in Evaluation Research", Evaluation Practice 9(4) pp.5-24.

Shapiro, J.Z. (1984) "Social Justice and Educational Evaluation: Normative Implications of Alternative Criteria for Program Assessment", Education Theory 34(2) pp.137-55.

Shapiro, J.Z. (1985) "Where We Are and Where We Need to Go" Educational Evaluation and Policy Analysis 7(3) pp.245-48.

Shapiro, J.Z. (1987) "Large-Scale Evaluation on a Limited Budget: The Partnership Experience", New Directions for Program Evaluation (J. Nowakouiski ed.) 36, pp.53-62.

Shapiro, J.Z. (1984a) "The Social Costs of Methodological Rigor: A Note on the Problems of Massive Attrition", Evaluation Review 8(5) pp.705-12.

Shapiro, J.Z. (1986) LSYOU (Louisiana State Youth Opportunities UNlimited) Project Evaluation ERIC ED 275815 CE045329 pp.1-64.