Searching Near, Far, And Wide: A Plan For Evaluatlon

Sylvia T. Johnson
Howard University

A Plan for Evaluation

Planning an evaluation for any major national program is a complex task. Often similarities in structure across program implementation in various sites serve as the basis for implementing traditional evaluation designs. If it is a service-oriented national action program, such as Women, Infants, and Children (WIC) or Headstart, there are certain parameters of input, as well as specific outcomes that can be measured and compared, even though specific projects have unique characteristics.

Many funded research programs have common parameters. The requests for proposals may have been structured to elicit examination of certain key constructs, methodologies, instrumentation, or populations, and these may provide the base for evaluation.

Educational research programs of the National Science Foundation (NSF) have goals that are primarily aimed toward expanding the envelope of scientific knowledge and being on the cutting edge of research. Such programs elicit a variety of proposals from researchers with considerably greater variety in terms of constructs, methodologies, and instrumentation than might typically be obtained. They also pose a more formidable challenge to the evaluator.

The Research in Teaching and Learning (RTL) program as well as other divisional programs present delivery models different from traditional school mathematics and science, and projects may vary in size, scope, and focus. Of course, there are intended effects of these programs. However, the variety of approaches and strategies employed, and the broad range of intended effects, spur the search for a method to examine and identify a number of different ways in which these programs may have left their marks—hence, the concept of footprints, left firmly, sufficiently protected from the elements, and molded well enough to be examined, understood, and replicated, and then converted into sturdy trails for the advancement of young learners of science and mathematics.

This paper presents an approach for developing an evaluation of programs composed of diverse projects. A general orientation to the task and the evaluation perspective employed is presented, followed by an overview of the one such diverse program, Research in Teaching and Learning (RTL). That program is then used as an example. Questions that an evaluation should address, and some ways of approaching them, are then presented. In the process of forming the questions, present and former program officers were interviewed. Included are suggestions prepared by a Research in Teaching and Learning Panel convened in the summer of 1992.

"Educational research programs of the NSF have goals that are primarily aimed toward expanding the envelope of scientific knowledge and being on the cutting edge of research."

The Evaluation Perspective

If one could examine a complex program of funded research from an all-knowing perspective, what could be seen? In developing a strategy or plan

"... it can be useful to examine it initially from this omniscient perspective ... to take an almost 'divine' perspective, if you will, and see where it leads you."

for evaluating a program of this type, it can be useful to examine it initially from this omniscient perspective, that is, to think of all the things that it would be great to know about it, even though they may be impossible to know—to take an almost "divine" perspective, if you will, and see where it leads you. The broad diversity of research activities funded, especially when that diversity is along so many dimensions—target populations, techniques, methodologies, etc.—further encourages this initial perspective in considering the evaluation task.

This omniscient perspective would go backward and forward in time to examine intention and planning, as well as long- and short-term outcomes. It would cut across all levels of researchers, participants, and other interested or not-so-interested parties. The outcomes would include those conventionally measured, and those virtually immeasurable. It would include the full range of unintended outcomes, both positive and negative, including those unknown and unknowable to the researcher and the ordinary human evaluator.

This perspective would go even further, though, in that it would discern what might have been. The solicitation, review, and selection of research for funding has many decision points, implicit and explicit. Suppose different directions had been taken in the identification of research projects for funding. Would there be important "Footprints" that are not currently in the picture?

Are there areas of desired footprints where we see more evidence of activity? What areas of possible effects show no effects? What footprints are missing?

In a sense, these are questions of ontology. To the logician, the question "what is there?" can be answered "everything," which while true, may not be especially informative, since the elements included may range from the universe to the empty set. Yet they are questions worth raising as a beginning point when the areas of possible effects are broad and diverse. A program officer also noted the need for an epistemological view in determining the extent and value of the "pay-off" from funded projects, because the created knowledge is invisible, and the extent of its utilization difficult to identify.

It would seem that this perspective calls for the evaluator to measure the immeasurable, observe the invisible, assess what might have happened if something else had been done, somewhere else, by someone else—a discouraging task, to say the least. In fact, the perspective being advocated here is meant to broaden the sensitivity, thinking, and powers of observation of the evaluator so that a more complete and useful appraisal of the program can be made. When one studies abstract art, or jazz music, or abstract mathematics, one begins to see, or hear, or conjecture more intensely, carefully, and ultimately, more clearly and with greater satisfaction and sense of thoroughness. When one is observing and enjoying a woodland scene, one can see, appreciate, learn, and enjoy even more, albeit somewhat differently, under the guidance of a trained forester, field entomologist, or ornithologist.

The goal of this exercise is to become sufficiently open to experience, information, and ways of knowing so that in developing an evaluation design and examining the many aspects of a complex program one can identify the need to measure a wider range of constructs with more diverse (perhaps, but not always) but less quantitative measures.

As a result, one should begin to see more as one looks more and more carefully, understand the logic of what alternative implementations might have made sense, where they might have occurred, and who might have been the most appropriate persons to have done them.

Crucial to this perspective is an openness and acceptance of alternative ways of knowing (Gordon, 1992), a willingness to question broadly a range of sources, and the time, interest, and wherewithal for sustained observation. Some vital occurrences do not occur often, and only the persistent may receive the reward of witnessing them. Scientific knowledge emerges from careful observation, yet sometimes dependence on conventional documentation limits discovery. While in no way should we expect to discard all of what we know about sound evaluation practice, neither do we limit our observations to conventional models. An approach that is open to receiving data from alternative sources is more scientific, not less so, because it means more careful observation and attending to alternative outcomes (**y**'s from a given **x**, and receptivity to alternate **x**'s as explanations for a given **y**).

This open and questioning attitude means, for starters, the questioning of oneself as evaluator, and repeating this among the evaluation team. It then means that more than the usual suspects are interrogated, and actually listened to.

Conventional methodology, in terms of examining specific projects, describing their inputs, and examining results of outcome measures does have a place in such an approach. In fact, the evaluation could be conceived as having three tiers: the first based on more conventional outcome data from projects; the second focusing on the footprints of the program in terms of impact and utilization, and the third looking for untouched areas, or the absence of footprints. For tiers two and three, the loci of the footprints (or non-footprints) are developed through a series of questions that examine effects on the program, on other research, on practice, and on other institutions.

In the following section, an overview of the RTL program is presented. From the perspective discussed here, a set of possible initial questions is raised. These questions, of course, would be supplemented by others as the thinking continues, and as initial data are collected.

Program Overview

The RTL program was begun in 1984 to support new discoveries about how individuals and groups learn, teach, and work more effectively in complex, changing environments. To this end, the program supports basic and applied research on factors that underlie the teaching and learning of mathematics, science, and technology at all levels. The program aims to support cutting-edge research, and has current priorities to look at the following issues.

1. How students learn complex concepts in science and mathematics.
2. How advances in knowledge of mathematical modeling link to learning complex concepts in science.
3. How teachers' subject-matter knowledge and competencies affect student learning.
4. How teachers learn to become inquiring practitioners and active researchers and how they learn to apply that knowledge in their classrooms.

"Scientific knowledge emerges from careful observation, yet sometimes dependence on conventional documentation limits discovery."

The impact of RTL studies on educational decision making by parents, teachers, administrators, scientists, policy makers, and curriculum developers at all levels regarding student literacy in science, math, and technology of knowledge is an important concern. Program staff also try to incorporate this generated knowledge into teaching methods and educational products that have direct usefulness in educational programs.

The program is aimed at teaching and learning by persons of all ages in formal school settings from elementary school through college, and informal personal and public settings. Accordingly, projects are conducted in broadly differing environments—classrooms, labs, homes, museums, conference halls—with a variety of methods and techniques from the cutting edge of work in these areas. About a quarter of the projects seek to improve understanding of special needs of learners and teachers traditionally underrepresented in scientific careers or whose needs for scientific literacy have not been met. These include women, African Americans, Hispanics, Native Americans, the physically or cognitively disabled, the gifted and talented, and learners whose native language is not English.

Another quarter of the projects examine motivational, attitudinal, or affective factors in learning and teaching with a focus on family, social content, cross-cultural differences, teacher beliefs, or classroom interactions.

The major goal of the RTL program is to generate a knowledge base that informs the national effort to reform mathematics and science education. Within this goal, activities of the program are aimed at achieving the following objectives:

- Supporting research on teaching and learning specific knowledge domains (chemistry, physics, mathematics, biology, computer science, etc.) at both the precollege and college levels, placing strong emphasis on establishing the content and sequence of learning that can be most effective in developing science and mathematics literacy and problem-solving skills.

- Building a coherent and comprehensive base of knowledge on learning and teaching in mathematics, science, and technology to meet future and current needs of decision makers, practitioners, and the research community.

- Encouraging research that will inform the reconceptualization of measures of performance and provide alternative methods for assessing student learning.

- Seeking research projects on the effects and significance of the nature and quality of laboratory experiences at all levels.

- Exploring factors that may influence interest, participation, and achievement in science and mathematics; development of motivation and curiosity; and the making of and persistence in, curricular and career choices at various student ages and educational levels, with a special emphasis on factors that influence underrepresented groups in their choices of course of study.

- Initiating an emphasis on direct teacher involvement in educational research so that questions arising out of classroom practice will more effectively inform the perspectives, methodologies, and findings of such research.

- Helping assure the application of research findings by teachers, teacher educators, policy-making educational administrators, parents, and other researchers.

What Questions Should Guide This Work, and How Will They Be Answered?

The broad program goal—generating a knowledge base that informs the national effort to reform mathematics and science education—along with the implementation objectives, provides the framework to generate questions. Other questions may be generated by interactions between objectives.

Impact and utilization are clear watchwords of the RTL program. The evaluation design should be centered on these terms, but with two thrusts. The first is a more traditional set of questions, using data conventionally explored in such investigations. These include the following:

- What publications were generated by the study?
- What awards were received by RTL researchers for publications based on RTL projects?
- How many undergraduate and graduate students have been supported by RTL-funded programs? What indices are available on their productivity?
- What conference and seminar presentations have resulted from RTL projects?

The second impact and utilization thrust is a less traditional one, and involves the utility of new knowledge and its effect on practice. Here we are examining impacts from the level of actual classroom practice, through teacher change, to effects on policy formulation in the education and political communities. The impacts of interest are often connected to studies with a rather traditional experimental sort of format, but the evaluation plan should relate to impact of new knowledge on practice. Such a format is the following:

- How do people (children, teachers, etc.) come to know and understand [concept, procedure, or configuration] **y**? How does [software, metacognition, instructional strategy] **x** help this process?

The evaluation plan then needs to examine questions of this sort in terms of the entire program.

- What are the influences on classroom practice, in terms of differences in what goes on in the instructional process, and in outcomes for learners? The outcomes should not be confined to problem solving and laboratory skills, although these are certainly of interest. They should include attitudes toward science and mathematics, interest in pursuing a mathematics or science career, interest in electives in science and math, and math and science interest and inquiry orientation, such as use of evidence in decision making, visiting science exhibits and museums, reading popular science periodicals, etc.
- What effects have RTL projects had on the research and develop-

ment community in terms of changes or developments in text materials, computer software, and teacher education? Curriculum material could be surveyed for answers here.

- What has been the effect of the emphasis on videotape technology in RTL projects? Has it had an effect on teaching practice? Are there steps needed to broaden the effects? What are specific examples of high impact? Can these be broadened? Richness of evidence of instructional value and quality is often applied to videotape. What evidence supports this, and does it show impact in practice?

- A reported impact on teachers has been that the research on children's thinking and mathematical understanding has empowered teachers; that is, as they have found that children have "incredible ideas," significant teacher enhancement has been reported. What documentation supports these incidental teacher effects, from studies which actually focus on children's thinking? What techniques would make this effect more broadly experienced?

"The scope of 'non-footprints' or at aleast fewer and fading ones, is an area of concern."

There is an issue of what is not being done, or is being done insufficiently. The scope of "non-footprints" or at least fewer and fading ones, is an area of concern. Staff have indicated a need to get new players into the research community, and have pointed out the problem of the aging academic cadre. Many research settings are not where problems in our schools are located. Think tanks opt for less harsh surroundings, as do most universities. But should RTL focus more on a broader base of populations? The Eisenhower Project of The Department of Education is important in expanding this direction, but there is certainly need and room for more. Yet there is still the need for basic research, and RTL is one of the few sources funding this work.

- What is the evidence of impact on utilization of new knowledge on mathematics and science teaching on what is actually going on in classrooms, as well as student outcomes in low-income communities, particularly those in schools serving African American, Hispanic, and Native American children?

- The program overview indicates that one-fourth of all projects were aimed at these students. Did these studies involve sufficient resources to maximize impact?

- Are program solicitations distributed to institutions that would be likely to carry out RTL work in inner city settings? Are workshops and professional group information sessions provided to encourage participation?

- What outreach activities related specifically to RTL studies are directed toward newer and nontraditional professionals? To what extent are they involved in panels and related activities?

- Some research centers have been very successful in RTL projects. They have been consistently funded, and their work has resulted in extensive publications, research-related projects, and the development of young scholars. What factors are related to the success of these projects? In what ways can their impact be broadened?

Collaboration is an important objective of RTL projects. It is encouraged within individual research projects as well as across the program. It is cited by staff as a primary objective of all projects. Another objective involves teachers as researchers, both to develop their inquiry and teaching skills and to impact students.

- Have collaboration and involvement of teachers as researchers been used extensively in inner city schools in RTL research projects?

- Has collaboration encouraged new researchers to seek RTL funding?

- How do teachers who have participated in RTL programs feel about collaboration?

- How has collaboration encouraged activity within the scientific community and between the science and math communities?

Here the work of recent years on standards in math and science, and the importance of these for assessments and teaching should be stressed.

In general, the questions above relate to effects on practice, the profession, the development of new research, and other institutions. Several sources of data are implied directly from the questions.

Other types of evidence and methods of obtaining them are found in the report of a 1992 Research in Teaching and Learning Panel. The panel suggested that RTL go back to the planning for the development of the RTL program that occurred in 1977-78, and engage in the following activities:

- Look at how RTL-funded research has influenced research reported at professional meetings.

- Have an independent group evaluate the quality of reviews, both supported and nonsupported, and how the proposers have reacted to them.

- Develop a genealogy to assess the impact of NSF-funded projects on people, i.e., the number of researchers whose initial work emerged out of working on NSF projects as undergraduates, graduate students, postdocs, consultants, etc., and how they developed as professionals.

- Assess the number of people recruited to the field as a result of NSF-funded projects.

- Document the impact of the program by asking people about their impetus into research in teaching and learning (autobiographies).

- Look at comprehensive reports that have reviewed projects funded by RTL.

- Assess the number and quality of journals that have been created as a consequence of the program.

- Assess the research agendas and their outcomes that have emerged from NSF-sponsored conferences.

- Look at PLATO, which has been a hothouse for future developments.

- Provide a snapshot of the people who have served on RTL panels.

"It should be noted that the time and resource base available to the evaluator is an essential consideration."

- Look at research reports from the American Educational Research Association (AERA), National Council of Teachers of Mathematics (NCTM), etc., to assess the number or percentage emerging from NSF funding.

- Look at mathematics and science educators who have broadened their views as a result of interactions with people outside the field, that is, look at the people who have served as consultants and on teams of the projects.

- Assess the time and efficiency of the program relative to NSF structure.

- Do a contrasting analysis of the mathematics and science communities.

- Look at applied journals for both authorship and citations, e.g., Physics Teacher, Science Teacher, and Mathematics Teacher.

- Assess the movement of people into other areas.

- Assess how many proposals in Teacher Preparation and Teacher Enhancement programs and the Instructional Materials Development program build on RTL-sponsored research.

- Assess the extent to which research is blended with practice.

- Look at the research discussed at NCTM conferences.

- Look at how RTL has affected programs at other foundations.

- Assess the impact of research on frameworks and standards.

- Conduct an ERIC keyword search.

- Look at all the regional laboratories and assess what they are disseminating.

- Look at the impact of the research and teaching methods that have been developed as a result of RTL-funded projects.

The questions and data collection sources and procedures above provide a beginning framework for the examination of RTL projects and the impact and utilization of new learning and discoveries. They provide multiple ways of knowing more about the program and its consequences. A parallel examination should provide similarly for other diverse programs of funded research.

It should be noted that the time and resource base available to the evaluator are essential considerations. The evaluator is not in sole control of the evaluation. The approach advocated here requires that the funding source allows for sufficient resources, time, and access to allow the kinds of things to happen that enrich the quality of the data and the evaluation report. If external constraints do not allow for this activity, the evaluation may be severely limited, despite all the openness in attitude conceivable.

Finally, the fact that an evaluation developed using these guidelines focuses on a broader range of evidence than is often considered should in no way be interpreted as minimizing the importance of rigor. Nontraditional does not mean sloppy, nor does it provide an exception to careful, intensive work. In fact, doing such work well is often more difficult and time consuming than working with "hard" data. Rhetoric is no substitute for data, but good science means careful observation and the accumulation of evidence from different sources, carefully and responsibly reported. Nor should a nontraditional label serve as a rationalizing shield for those using traditional statistics poorly, and claiming that their work is not accepted because they "aren't hung up on a lot of statistics." Nontraditional evaluation does not depend on magic: just on science thoughtfully conceived, coherently organized, and clearly reported.

References

Frechtling, J. 1993. *Research in teaching and learning* (background paper).

Quine, W.V.O. 1963. *From a logical point of view*. New York: Harper and Row.

Stake, R. 1975. *Evaluating the arts in education*. Columbus, Ohio: Charles E. Merrill.

Worthen, B.R., and Sanders, J.R. 1987. *Educational evaluation.* New York: Longman.

Gordon, G.W., and Bhattachanyya, M. 1992. Human diversity, cultural hegemony, and the integrity of the canon. In *Journal of Negro education*, 61: 405-418.