Communicating The Value Of The National Science Foundation's Contributions To Research And Innovative Technical Applications For Mathematics And Science Education

Norman L. Webb
Wisconsin Center for Education Research
University of Wisconsin

Introduction

When evaluating National Science Foundation (NSF) programs that fund research and innovative technical applications in mathematics and science education, it is important to consider the main purposes of the evaluation. One well-established purpose calls for the evaluation to identify the effects of the program—on the profession, on other research, on practice, and on other institutions. But focusing attention on effects tends to direct attention away from the intended audience of the evaluation. Stated differently, the evaluation of NSF programs should not only identify the effects of programs, but the evaluation should communicate the value of those effects to a variety of audiences—the United States Congress, the mathematics and science education professions, the NSF administration, and the public. Indeed, the value of research and innovative technical applications is often greater than its immediate effect, and any evaluation that fails to communicate this value will fail to live up to its potential. Consequently, in the effort to design nontraditional approaches to evaluation that is presented in this paper, I argue that the determination of a program's value should be integrated with the communication of its value.

A full appreciation of promising approaches to the evaluation of NSF programs that fund research and innovation requires an understanding of several factors. The Research in Teaching and Learning Program, the Applications of Advanced Technologies Program, the Educational Indicators and Studies Program, and other NSF programs are complex. Each program has multiple goals, incorporates expectations that are not always clearly articulated, uses limited resources to solve large problems, and is required to be sufficiently flexible that it both responds to immediate concerns and prepares the Foundation for future needs. Given this complexity, a productive evaluation of these programs should draw upon knowledge of at least the following:

- The nature of research, innovative development, and research-driven enterprise;
- The long-term pay-offs of some kinds of research;
- The most promising lines of inquiry at any given time;
- The past record of established researchers;
- The need to nurture young researchers; and
- The relative importance of groups that have a special interest in the research.

The evaluation of NSF programs that fund research and innovation is further complicated by the nature of mathematics and science education. The teaching and the learning of mathematics and science are different. Each field has

"... the evaluation of NSF programs should not only identify the effects of programs, but the evaluation should commuicate the value of those effects to a variety of audiences ..."

different curriculum needs and traditions. Those who work in or interact with each field vary greatly in their interests, work, and demands that are placed on them. This observation applies with equal force to teachers, teacher educators, students, researchers, scientists, mathematicians, school administrators, and policy makers. Each field of mathematics and science education has its own community of scholars and researchers. Nonetheless, NSF programs must serve both fields and, at times, must even allocate resources among the researchers who work in both fields.

As a body of inquiry, evaluation itself adds to the complexity of determining the value of governmental research programs. Studying and evaluating an NSF program has political overtones and ramifications. In addition, amidst calls for public accountability for programs of this kind, the task of assigning a value to the work of the program may create some troubling paradoxes. Specifically, the evaluation of research that is carried out under a given program may validate the high quality of one set of research findings that run counter to the findings of other well-publicized and developed projects supported by the same agency. Further, because each NSF program funds a wide spectrum of projects, this situation could even occur within an individual program. Finally, the costs of evaluation also add to the complexity of determining the value of programs. The benefits of an evaluation to the program and the Foundation must be weighed against the expenses of conducting evaluation that can adequately deal with the multifaceted composition of the program. Since these kinds of factors are important practical constraints upon program evaluation, they should be considered when designing and selecting models for the evaluation of NSF programs.

In this paper I have tried to speak to some of these concerns. However, a full explication of these factors and their relationships to evaluation would require a major document. My intent here is to offer sufficient explanation that the rationale for each nontraditional approach to evaluation is made clear.

In brief, I recommend a series of evaluation studies, since the varied characteristics of the studies best accommodate the variety of goals that a program can have.

- One recommended study, a retrograde analysis, considers how funded projects have built on and used findings from previous projects that were funded by a program. The retrograde analysis is designed to communicate the integrity of the programs and to show how funded research and projects have built on each other to develop a body of knowledge that is being applied to science and mathematics education.

- A second proposed study, a video documentary, is designed to use visual images to communicate the findings and innovations that have been generated through NSF programs and to elucidate their value to educational practice.

- The purpose of a third proposed study, a research community culture analysis, is to communicate the richness and productivity of the community of researchers that has evolved, at least partly, because of funding that it has received from NSF. A significant number of people have served on NSF-funded projects and have gained knowledge and experience while working on those projects. The work and expertise of these researchers and others extend beyond the boundaries of the work that they have performed for the NSF. An analysis of this community can reveal some of the extended effects of NSF programs.

- The fourth proposed study, generalizability analysis, is an attempt to attend to the spectrum of impacts that NSF programs can have. The analysis would use sampling techniques and large-scale instruments to produce information about results from a collection of funded projects, and the analysis would attempt to identify the impact of those studies upon likely users.

The body of this paper begins with a brief description of one NSF program, Research in Teaching and Learning (RTL), to exemplify the complexity of a funding program and the wide variety of projects that are funded. The other programs that are pertinent to this study, such as the Applications of Advanced Technologies Program and the Educational Indicators and Studies Program, have comparable characteristics and are equally diverse. The description of the RTL program is followed by a discussion of the diversity of research in education. This discussion is followed by statements of specific evaluation questions that are central to this kind of undertaking, and by a brief enumeration of issues and pitfalls that are likely to arise in the evaluation of NSF programs. The paper concludes with an outline of four promising approaches to program evaluation that would communicate the value of NSF programs to the most important audiences that could use the results of this kind of evaluation.

Brief Description of the Research in Teaching and Learning Program

Overcoming conceptual difficulties in science; generating more and better mathematical discourse in elementary classrooms; building models of student achievement in science and mathematics; identifying the theoretical and national policy implications of the persistence of high-ability minority youth in college mathematics, science, engineering, premedicine, and predentistry programs; and assessing changes in home processes related to children's interest and proficiency in mathematics as they are affected by a program designed to help parents to be more active in their children's mathematics learning: these are only a few examples of the 187 new and continuing projects that were supported by the Research in Teaching and Learning Program during the 1987-91 period. Over these 5 years, grants totaled $23.45 million—not including the funding of 26 projects that were shared jointly with other NSF programs between 1987 and 1990. Significantly, in terms of the numbers of projects funded, the greatest concentration of awards was in the field of mathematics, followed by physics, general science, interdisciplinary area, biology, chemistry, and astronomy.

"Research in Teaching and Learning... seeks to support new discoveries about how individuals and groups learn, teach, and work more effectively in complex, changing environments."

Goals of the Program

Research in Teaching and Learning—a program in the Division of Research, Evaluation and Dissemination—seeks to support new discoveries about how individuals and groups learn, teach, and work more effectively in complex, changing environments. RTL supports basic and applied research to answer questions about the teaching and learning of mathematics, science, and technology at all levels. Findings from this research are to inform those who are active and interested in education and its reform. Policy makers, teachers, teacher educators, curriculum developers, parents, and researchers are among the people who compose the intended audience for the research output and findings that appear in reports, videos, computer software, laboratory activities, and instructional materials. Although RTL has been supporting research since 1984, its current priorities are to advance our understanding of the following:

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

The goal of the RTL program is to generate a knowledge base that informs the national movement to reform mathematics and science education. To attain this goal, the program has specific objectives.

- First, the program seeks to establish the content and sequence of learning that can be most effective in developing science and mathematics literacy and problem-solving skills.
- Second, the program endeavors to meet the current and future needs of decision makers and other people who perform critical roles in education and research by building a coherent and comprehensive base of knowledge of learning and teaching in mathematics, science, and technology.
- Third, RTL seeks to produce research that will inform the reconceptualization of performance measures and that will develop alternative methods for assessing student learning.
- Fourth, the program is to study the significance of the nature and quality of laboratory experiences and determine their effects.
- Fifth, RTL is to explore factors—especially those influencing underrepresented groups—that empower students to participate and achieve in science and mathematics and to develop a positive disposition toward these fields of study and work.
- Sixth, the program seeks to engage teachers in education research, as a strategy to help make findings become more closely attuned to classroom reality.
- Finally, RTL's seventh objective is to assure that research findings are applied by members of the education community—teachers, teacher educators, policy makers, educational administrators, parents, and other researchers.

Range of Projects Funded by the Program

Projects supported by the Research in Teaching and Learning Program vary in their purposes, methods, age levels of student populations, and subject matter. As indicated by the nature of the projects that were cited at the beginning of this paper, goals of projects can range from addressing policy issues and providing information for policy decisions to very specific learning problems. The program uses five categories to group and describe the range of its projects: setting the research agenda, research in teacher enhancement, research on student learning, curriculum research, and cross-cultural research.

RTL involvement in setting the research agenda includes supporting major conferences, reports, and publications within the research community. Recent funding has been directed toward research projects that advance current efforts to reform mathematics and science education. For example,

- The “NCTM Research Catalyst Conferences” had six groups of researchers, each of which involved two mentor researchers who met with less experienced researchers, to design and encourage research critical to the implementation of the National Council of Teachers of Mathematics (NCTM) Curriculum and Evaluation Standards for School Mathematics.

- Another RTL-funded project prepared the aptly titled report “Establishing the Research Agenda: The Critical Issues of Science Curriculum Reform.” This report was discussed at national meetings and published in the Journal for *Research in Science Teaching.*

Other funded projects have helped to define the research agenda in education by summarizing key research findings and by examining ways that findings can be communicated to practitioners.

Funded research in teacher enhancement targets the teaching process and reveals ways that student learning of mathematics and science can be expanded.

- The Cognitively Guided Instruction Project, directed by Elizabeth Fennema and Thomas Carpenter at the University of Wisconsin-Madison, and funded jointly with the Division of Teacher Preparation and Enhancement, produced research-based materials and strategies for inservice and preservice teachers to be more effective by using knowledge about student thinking to make instructional decisions.

- Another example of funded research in teacher enhancement is a school-based research project that is run cooperatively by the University of Maryland and the Montgomery County Public Schools in Maryland. Project Impact (Increasing the Mathematical Power of All Children and Teachers) strives to enhance student understanding of mathematics through summer inservice programs for teachers of minority children.

In the course of these programs, teachers study pedagogical content knowledge, mathematical content knowledge, and their beliefs. Teachers use the opportunity to examine and develop

“... goals of projects can range from addressing policy issues and providing information for policy decisions to very specific learning problems.”

instructional activities that foster mathematical understanding and problem solving. Evaluation is ongoing in studying the implementation of the summer inservice goals and in a multiyear impact evaluation of the effects of the inservice programs on student learning and teacher beliefs and practices.

Research on student learning embraces projects that focus on student cognition, concept learning, problem solving, and the knowledge that students bring to the formal educational setting.

- In science, funded projects are devising and studying new ways to help students learn such traditionally difficult concepts as force, motion, gravity, harmonic motion, and the adaptation and natural selection mechanisms that underlie biological evolution.

- In mathematics, funded projects focus on topics that range from early number concepts through multiplication, estimation, pre-algebra and algebra, geometry, calculus, probability and statistics, and abstract algebra at the college level.

Curriculum research includes projects that endeavor to inform instructional materials development. Research focuses on topics from the school and college curriculum and is designed to foster curricular and instructional innovations.

- In science, research on topics in physics, chemistry, and biology help to structure instructional materials.

- In mathematics, other studies focus on Logo geometry for elementary schools, mathematical modeling and exponential functions for high schools, and calculus concepts and computers for courses at the college level.

Finally, in funding cross-cultural research, RTL intends to raise the expectations of educators concerning student achievement and classroom practices by studying practices and results from other countries.

- The work of Harold Stevenson and others on Japanese, Chinese, and American students has been highly acclaimed and widely published in both the scientific and popular press. These researchers have articulated their objectives in the following terms, "the goal of this research is to increase understanding of prior and contemporary influences on achievement in mathematics so that effective suggestions may be made for the improvement of mathematics education in the United States" (*NSF Summary of Awards, Research in Teaching and Learning, Fiscal Years 1987-1990* [hereinafter NSF 1987-90] 80).

- Other research projects in this category seek to maintain and enhance the database of the IEA Second International Mathematics Study and provide American educators access to research monographs that have been published exclusively in the former Soviet Union.

The five categories of research projects that are supported by the Research in Teaching and Learning program embrace diverse projects. The objectives of the projects that are included in these categories range from very broad issues of reform and international perspectives to very specific concerns in concept learning, teaching practice, and materials development. Moreover, the range of project goals within any given RTL funding category is very extensive and broadens, rather than concentrates, the diversity of research endeavors. For example, under the category of research on student learning, some studies use students' mathematical errors as a springboard to critical thinking (NSF 1987-90, 6); another project focuses on systems of concepts in multiplicative structures; a third studies the cognitive processes that are involved in understanding and using scientific diagrams; and still another project is attempting to facilitate the process by which students learn to connect real-world phenomena with scientific representations of the phenomena. This variety in projects is also evident in the educational level that serves as the focus of funded research, as indicated below.

- Approximately 29 percent of the projects during 1987-91 concerned the elementary level of education;
- Fifteen percent concerned the middle school level;
- Thirty-three percent concerned the secondary level;
- Eighteen percent concerned the undergraduate level, and
- Nineteen percent were not related to any single grade level, since some projects treated more than one grade level.

Another indicator of the diversity of these projects is the fact that over 300 key words and phrases are listed in the index of the 1987-90 RTL summary report. In brief, NSF's program of Research in Teaching and Learning appears to seek broad coverage over concentration in the projects that it supports, since RTL addresses learning and teaching by people of all ages, and since RTL tries to provide information for decision making by a range of people, including parents, teachers, administrators, scientists, policy makers, and curriculum developers.

The Practical Nature of Research in Education

Educational research incorporates many kinds of inquiry and is not limited to a particular mode of investigation. The objectives of educational research can range from efforts to understand the learning process to the gathering of information that is intended to improve decision making. Borg and Gall (1983) have classified educational research into four typologies that differ according to the following characteristics of the research: topic, purpose, hypothesis testing, and basic versus applied research.

Topic describes the phenomena investigated, such as learning process, cognitive abilities, and teaching methods. Purpose addresses whether the research attempts to describe or characterize a group of phenomena, or whether it tries to reveal relationships among variables. Hypothesis testing research involves studies based on some prior theory or findings that are used to confirm or reject conjectures. Basic versus applied research distinguishes between research that focuses on understanding fundamental structures and processes (basic research) and research that focuses on structures and processes as they appear in educational practice (applied research).

"Research can benefit the development of curriculum materials by indicating what works and what does not work."

Research to Inform the Practice of Teaching and Learning

Although the nature of educational research is varied, education is a practical field that continually requires teachers, administrators, supervisors, and others to make decisions that have cumulative influences on the lives of students. Research that facilitates decision making, that provides guidelines to help reduce the complexity of educational content and instructional practices and materials, or that provides answers to questions that arise repeatedly has enormous potential for teachers and others, provided findings are put in a useable form. For example, knowing that 6-year-old students enter first grade with thinking strategies that are useful when solving mathematics word problems that have generally been presented to older students (Carpenter and Moser 1983) is a powerful finding that could help first grade teachers to work effectively with their students.

The curriculum, the goals, objectives, and instructional materials that are necessary to achieve desired outcomes, is a dominant force in determining what is taught in classrooms in this Nation. Research can benefit the development of curriculum materials by indicating what works and what does not work. Systematic feedback on draft versions of instructional and curricular materials can be critically important to curriculum developers who are writing materials for use in classrooms.

Research to Lead Reform

The relationship of research to education reform often incorporates an important bifurcation: research can prompt reform; or research can be a response to reform. To cite a significant example, the NCTM Standards were written by people from the research community and by other mathematics educators in the profession who were very knowledgeable about research findings. This knowledge—in addition to collective, accumulated experience in teaching and producing effective curriculum materials—was very valuable in the preparation of the NCTM Standards that have served as a driving force in current efforts to reform mathematics education in this country. Furthermore, the NCTM Standards went beyond existing, verified knowledge and established new expectations regarding the nature and extent of mathematics that all K-12 students should experience. The Standards also presented content topics (e.g., discrete mathematics) for which there were very few available curriculum materials. The Standards then set an agenda for additional research that would be needed to effect the vision that the Standards offered to the community of mathematics educators and researchers. In this manner, research can do more than add fuel to the fervor for reform by helping to ignite the flame and by adding tinder that will keep the flame going.

Research to Develop and Confirm Theories

Theory building, theory verification, and model building have been applied to education and have been an application of research. If we share the view of Kaplan that "a theory is a way of making sense of a disturbing situation so as to allow us most effectively to bring to bear our repertoire of habits" (1964), and that a model is "the embodiment of a structural analogy" (1964), then we can see that theories and models are useful in providing a language for communication and in making predictions. Indeed, with well-developed theories and models, predictions can be very precise. Piaget's theory of the development of intellectual capacity in children, and its focus on their attempts to structure their world and give it meaning, fostered a large body of research. Carroll's model of learning (1963) that depicted learning as a function of prior knowledge, perseverance, opportunity to learn, and other variables, was instrumental in the mastery learning movement and was used to design research to verify that model under different conditions. Because education is complex and involves many variables, educational theories and models are difficult to develop, but successive iterations in the development of these theories and models help to define research questions more precisely and productively, and link individual research studies to other bodies of organized inquiry.

Research to Explain Outcomes and Practice

A common use of research in education is to describe outcomes, practices, and conditions. Teachers who are isolated in their classrooms can benefit from descriptive studies that reflect on the practices of others. Such studies provide confirmation for a teacher who rarely has the opportunity to observe other teachers in their classrooms or to consider variations in teaching practices. National and international studies that describe the achievement level of large groups of students, or the achievement differentials by different groups of students, are helpful for policy makers when they review policies and allocate resources.

Education is notorious for borrowing direction and methods from many other fields, such as psychology, the natural sciences, anthropology, and linguistics. Educational research is no different and has applied a variety of methods to study questions that bear on the field. The range of methods includes ethnography (anthropology), computer simulations and models (computer science), case studies (medicine and sociology), statistical analyses(statistics), cost-benefit analyses (economics), and policy and historical analyses (political science and history). These methods of inquiry have an impact on the ways that researchers interact with their findings, and can reveal different information concerning the same phenomena. In light of the large number of variables, factors, and complexities that arise in most educational research, multiple methods of research are necessary if we are to begin to identify and understand the important variables and relationships among variables that exist in education.

Research in Science and Mathematics Education

The nature of research on teaching and learning in science education and in mathematics education is defined by a multitude of factors. In a certain sense these fields are very young. The bodies of knowledge, methodologies, and traditions that they draw upon are continually under development. Moreover, both fields are greatly influenced by the content areas of mathematics and science, and many researchers have been trained in those disciplines. In addition, the education of students in these content areas requires attention to psychology, learning theories, and educational foundations. Because education is so diverse, researchers in both science education and mathematics education have drawn upon many methodologies to study teaching, learning, curriculum development, and policies. The emerging technologies and their applications to education have required mathematics and science education researchers to expand their knowledge to understand these new and changing forms that have the potential to change drastically the teaching and learning of mathematics and science. Given these varied sources and methods in education, research on teaching and learning in mathematics and science education calls for corresponding variety in the approaches that are used to conduct research and maintain contact with research advances. This compounding of complex educational methods and research approaches often makes it difficult to understand the research, and to identify and communicate the value of the research.

"The communication of the value of the programs requires depicting what the programs have done, what their main effects are, and how these effects have been applied to practice."

Evaluation Questions

The questions that are to be answered in the course of evaluating such NSF programs as the RTL program should be structured by the purpose of the evaluation. As argued at the beginning of this paper, the central purpose of NSF program evaluation is the communication of the value of NSF programs that fund research and innovative technical applications in mathematics and science education. Communication is constructing knowledge. The acts of writing, speaking, reading, and listening require building on existing knowledge, making decisions, analyzing information, and drawing conclusions. The act of communicating the value of NSF also entails constructing the value of the programs by focusing on what is important, analyzing information, and drawing conclusions. The communication of the value of the programs requires depicting what the programs have done, what their main effects are, and how these effects have been applied to practice. But the communication process also attends to an audience and sends a message. As a central purpose for evaluation, the communication of the value of programs combines the substance of the message with the message itself.

Clearly, additional purposes for an evaluation of the effects of an NSF program can be phrased in other ways. One purpose could be to ascertain the accomplishments of the RTL program and the impact of these accomplishments on instruction and learning in mathematics and science in the United States. Two other purposes could be served by the evaluation: the undertaking can gather information targeted to strengthen the program, so that it will be more effective in achieving its goals; and the evaluation can reflect upon the goals of the RTL program. In reflecting on the goals of the program, attention would need to be given to their relationships with the goals of other programs, so that a clear view of the correspondence among goals can be obtained, and so that the future needs of mathematics and science education over the next 5, 10, and 15 years can be defined. In brief, the evaluation of the program needs to be specific enough that it can be accomplished, but it needs to be general enough that it will provide confirmation, direction, and a rethinking of procedures. Focusing on the communication of the value of the program can meet these criteria.

Sample Questions

In communicating the value of NSF programs, there are at least six important questions that the evaluation should strive to answer.

1. What research findings and information have been produced by individual projects and by the collectivity of projects that have been supported by the RTL program?

In thinking about this issue, it is useful to decompose the question into its components by employing the two-by

Exhibit 1

Four questions for an evaluation of the Research in Teaching and Learning program, structured by the information that is now known as a result of the research and by the applications of those research findings.

Research Results	Applications: Yes	Applications: No
Know	What findings and information have been produced that have successfully solved a problem or fulfilled a need?	What findings and information have been produced that have not been applied to solve an important problem or fulfill a need?
Do Not Know	What critical problems or needs have not been resolved or refined by research findings and information?	What negative or poor applications have filled the gap in the absence of solid research findings and information?

two matrix that is depicted in Exhibit 1. One dimension represents the information and findings that have been produced by RTL projects. This "research" dimension can be divided into two categories—what we know and what we do not know. The second dimension represents the application of research to existing problems. This "application" dimension can also be divided into two categories—what research has been applied and what has not been applied. This simple matrix helps to generate four classifications of questions that should be answered by the program evaluation.

1a. What findings and information have been produced that have successfully solved a problem or fulfilled a need?

The responses to this question will be the success stories of the program. Projects that have been successful in gaining results and in having these results applied to the solution of important problems will provide strong evidence about the impact of the program. An important part of the answer to this question lies in the identification of problems and needs and in demonstrations of the ways that funded research provided solutions to the problems or met the needs. In addition, it is critically important that the question and consequent answers focus on significant problems. For example, helping elementary teachers to learn how to build on student thinking in their teaching is more significant than deciding between the use of vertical addition or horizontal addition.

"The evaluation should determine what the program has not done in areas where information and research results would be useful."

Clearly, assigning importance to problems is a value judgment, and that reality should be considered in the design of any evaluation.

1b. What critical problems or needs have not been resolved or refined by research findings and information?

The evaluation should determine what the program has not done in areas where information and research results would be useful. Some explanation of why research has not been successful in resolving—or at least, in refining—important problems will need to be incorporated into the answers to this question. There may be many important reasons why research findings are not available. Possible explanations might include the following: research may have been tried, but findings may have been inconsistent; funds may not have been available to support the needed research; the research may not have been concentrated in the manner that would have been most likely to resolve the problem; and insufficient time or resources may have been allocated to solve the problem.

1c. What findings and information have been produced that have not been applied to solve an important problem or fulfill a need?

In any research program, some efforts will not produce the intended results or will not be productive. Alternately, some research may not address questions that are as important as other research. One would hope that there would be a minimum of such research that is supported by the RTL program. However, a program without any such efforts is probably insufficiently aggressive in advancing knowledge in a given area. Still other research will address basic questions whose answers do not have any immediate applications. For example, some psychological research in the learning of nonsense syllables is basic and lacks direct classroom applications. A complete evaluation of the RTL program would need to identify research efforts and findings that have not been applied and would need to assign some value to these efforts, since they may have made a significant contribution to a body of knowledge and may be an important outcome of the program.

1d. What negative or poor applications have filled the gap in the absence of solid research findings and information?

Any program that supports research will have to decide between the research that it will fund and the research that it will not fund. In some instances, important educational questions will arise, and no information from research may be available to help respond to those questions. The absence of this information may suggest that the program has failed to anticipate the issues that will arise in the future. In that event, practitioners will have to use the best information that is available to them. In some cases, the information that is available or the practices that are current may be relatively unsuccessful or may even produce poor results because the needed information has not been produced. For example, some feel that an extended use of mathematics worksheets with young children can result in rote learning and the development of a very mechanistic view of mathematics. Without research findings that refute this practice, some teachers will continue to have a worksheet-based mathematics classroom. Consequently, the evaluation of the RTL program should at least acknowledge the kinds of research that have not been funded and should consider the implications—both positive and negative—of the decisions not to fund certain research.

In addition to evaluation questions that focus on the application of research findings, there are five other questions that should be considered.

2. How has the RTL program contributed to the development of a community of researchers who serve as resources for the education system?

3. How have findings and information from the RTL program supported other program efforts, and how have the findings and information been used by other NSF programs, such as that in Instructional Materials Development?

4. How has the RTL program shaped and set the research agenda in mathematics and science education; and, more particularly, how has this agenda setting derived from provocative questions that have been formulated by the program and that have motivated large numbers of studies?

5. How have the RTL program and its funded projects built on findings from related research programs and fields, such as those in psychology and computer science, to ensure that supported research is relevant and does not duplicate work in other fields?

6. How have the operations and funding strategies of the RTL program served the program's goals?

Issues and Pitfalls in Evaluating the Research in Teaching and Learning Program

There are seven issues that are central to the design of program evaluations for the National Science Foundation.

- One issue concerns the unit of analysis for an evaluation. To show fully the extensiveness of the NSF program's accomplishments, whenever possible, the unit of analysis should be the program.

- Scale is a second issue. One major goal of the NSF is to improve the quality of the Nation's science, mathematics, engineering, and technology education. Trying to observe movement in the national system poses massive problems for the comprehensive evaluation of programs.

- A third issue is that the observation of important effects will depend somewhat on the time and duration of the research projects. Sometimes important systematic effects do not appear until years after the completion of a project. Also the research or project could have been worthy, but the project or research may not have been extended over an adequate period of time to produce observable systematic effects.

- A fourth issue is that change and the evidence of change is not uniformly apparent over the education system. The problem becomes one of locating the points at which change has been concentrated in the educational system, and of attributing the change to identifiable research and development projects.

"Sometimes important systematic effects do not appear until years after the completion of a project."

“...in studying NSF programs some consideration needs to be given to effects that go beyond those stated in projects’ proposals or final reports.”

- A fifth issue concerns the synergy of the research and education systems and how information flows between the two. Funded research may be of a high quality, but the dissemination of findings may be poorly implemented.

- A sixth issue in studying the impact of research on practice is that there may be conflicting forces that bear on the support of research and the application of research. What research has determined to be theoretically sound practice may confront current practice that is strongly embedded in tradition and values. Or, the recommended changes may be overwhelmingly expensive. Quality research cannot always be expected to find its way into practice.

- Finally, in any evaluation of research programs there are unintended outcomes that in many cases will be positive. This implies that in studying NSF programs some consideration needs to be given to effects that go beyond those stated in projects’ proposals or final reports.

Promising Approaches to Evaluation

Evaluating the impact of the NSF programs is complicated, as indicated earlier, by the great variety of projects that were funded under the programs, the range of age groups that were targeted by projects, the forces within education that retard the implementation of research findings, and the lack of concentration of results that can be brought to bear on the educational system in the United States. Tracking the effects of any one of the programs, such as the RTL program—on the profession, on other research, on practice, and on institutions—is further complicated by the many other influences that affect schools and education. Alternate approaches to evaluation are needed in order to reveal the levels of outcomes and the variety that exists among outcomes. In light of these considerations, some nontraditional approaches to evaluation can communicate to others the value of NSF programs that fund research and innovative technical applications for mathematics and science education. To help simplify references to the different programs, the four approaches to evaluation are described in the context of only one funding program, Research in Teaching and Learning. The approaches, however, could be applied to any of the other programs or to combination of programs.

Retrograde Analysis

One indicator of a research program’s value is its internal integrity: how research produced over the years builds upon research that was previously produced by the program. A program with internal integrity will develop a coherent body of knowledge with evident chains of inquiry. The value of the program, in this case, is the created body of knowledge that can be drawn upon by different people for multiple reasons. Strong chains of inquiry are more apt to lead to significant applications when ideas are highly developed, expended effort and resources have been concentrated, and findings have stood the test of time. Communicating the value of created bodies of knowledge becomes a problem of describing what the body of knowledge is, how it has evolved from the work of projects within the program, its importance, and its potential applications.

The study of the internal integrity of the RTL program and the bodies of knowledge that it has generated could be done by a team of three people—one evaluator, one science educator, and one mathematics educator. The principle charge to the evaluation team would be to analyze the relationships that exist among the findings of projects that have been supported over time. The central focus of the evaluation would be to document the relationships among the findings of the most successful projects and to establish the fact that projects have built on each other to form coherent bodies of knowledge. The work of other projects could be studied as appropriate or warranted. The most productive projects to begin this investigation can be identified from the amount of funding received, the visibility of the project, and the extensiveness of findings. The Cognitively Guided Instruction (CGI) project is one example of such a "star" project.

Instead of the usual approach to evaluation, which examines the progression from early studies to more recent studies, it would be useful to proceed in a retrograde manner, by examining the ways that more recent studies have relied upon and built upon a succession of earlier studies. Such retrograde analysis would examine relationships between funded projects by focusing upon the "generation" of the projects under consideration —by moving from the current research generation to research that was funded and conducted one, and two, and three or more generations earlier. In this approach to program evaluation, what is currently known from each of the "star" projects could be described by using information obtained in interviews of the project staff and others, by reviewing project documents and technical papers on findings and results, and by surveying other sources of information as appropriate. Then, one could analyze the research bases for the current findings and information, and the derivation of these research bases from research that was conducted one and more generations earlier. In this manner, a project genealogy would be produced (Webb, Shoen, and Whitehurst, 1993). Subsequently, the linkages between research generations would be used to identify the initial or formative ideas that underlie research over time. The intent in this approach to analysis is to establish a chain of inquiry linking the generations of projects, and to relate this chain to support from the RTL program or to the manner in which RTL has built upon support from other sources. Such an analysis has the potential to demonstrate the cumulative or building effect of research findings, the evolution of projects over time, the evolution of project staff thinking, and the productive use of RTL funding. The most likely chains to be revealed are ones that follow a researcher, group of researchers, or a topic of research. Theoretical mappings and idea tracings over time are possible outcomes.

Chains of inquiry and other findings from this analysis can be validated by direct evidence—a researcher reporting and showing evidence of a link to work of another project—or triangulation of evidence—confirming evidence received from different sources. The final product of this evaluation could be a report including both a narrative explanation of the linkages found and charts depicting the development of bodies of knowledge.

Video Documentation

A second evaluation approach to communicating the value of the RTL program builds on Marshall McLuhan's idea that the medium is the message. The central evaluation question focuses on the coherent messages about classroom practices and educational innova-

"Instead of the usual approach to evaluation, ... it would be useful to proceed in a retrograde manner, by examining the ways that more recent studies have relied upon and built upon a succession of earlier studies."

tions that can be gleaned from the program. The form of reporting findings from this investigation would be a video documentary. The process of creating the documentary will be, in and of itself, an evaluative investigation extended further by using the different elements available in video to communicate the findings. Video is a powerful medium for reporting to large and varied audiences. Video, as compared to text, has the advantage of communicating more clearly the visual materials that are produced by projects, new applications of technology, and the full range of diverse projects that form the program.

The preparation and production of the video RTL documentary would be the responsibility of an evaluation team consisting of an evaluator, mathematics and science educators, a producer, a script writer, and necessary production support staff. The time that any one person would spend on the evaluation would depend upon the extensiveness of the study and the role to be assumed. The process would begin by researching and analyzing the main messages that can be derived from the RTL program. Then, the selection and focusing process would identify the major theme or themes for the video, based on validated findings, what has been put into practice, what is visually exciting, and what is ongoing, exciting work that has the potential for change. Subsequently, an editorial board, consisting of NSF staff, researchers, and others, would critically analyze the themes and the work selected to create the video and to substantiate the selections of material. The evaluation team would need to have some autonomy to do the necessary research, prepare the script, and produce the video. Some written materials could be prepared in support of the video, but the video should be the main form of communication.

The actual story and the major themes of the documentary will be decided as part of the process of evaluation. Many possibilities exist.

- One is to report on actual classroom applications where practices have been directly influenced by RTL projects. A variation in focusing on classroom practices would depict the applications of research findings by making a composite of an ideal classroom for different grade ranges and content areas. Classroom composites could reveal in concrete terms the practical body of knowledge that has been generated by funded research. The classroom composites could consist of written and video scenarios of the RTL-influenced classrooms that depict teaching practices, student activities, and student learning.

- Another possibility for the story line of a documentary would be to take an issue, such as opportunity to learn, and show how RTL projects, such as the Second International Mathematics Study (SIMS), have advanced our understanding of that concept and how there are consequences that can be documented or anticipated from this advancement. For example, SIMS data indicated that opportunity to learn was strongly correlated with achievement, as has been supported by other studies. This can be a powerful message when thinking about "world class standards." A treatment of opportunity to learn could also lead to a timely analysis of equity, and to an analysis of differences in content and in presentation to various groups of students.

An evaluation of the value of RTL and other programs would grow out of the process of revealing the implications of what we know to be true and what we think is possibly true.

In addition to investigating major themes across the RTL programs and their applications to practice, the video development process can be used to reveal the questions that projects are pursuing and the substance of what is being learned. Many projects use video as a research tool to record student interviews and classroom interactions, and a video documentary could build on these video resources that communicate very well what has been developed. This could be accomplished by collecting video and other visual materials from projects, by abstracting depictions of new findings and applications, and by creating video episodes to present the major ideas. This process serves both as a means of evaluating the richness or weakness of findings and as a form of communicating and describing some of the RTL program's effects.

Other video techniques afford unique ways of communicating the range of findings, the scope of work, and the applications to practice. Some of these techniques are:

- Video interviews with researchers, teachers, and students;
- Voice-over segments that illustrate a new practice while the audience hears a teacher reflect on the practice;
- Montages that present a range of investigations through a sequence of music-accompanied images that are flashed on the screen; and
- Presentations of computer simulations, software demonstrations, or CD-ROM applications to explain the wide use of technology that is being supported by RTL.

The process of producing a video documentary using these and other techniques, along with presenting major themes and applications, requires grouping RTL work and findings into categories, deducing meaningful conclusions, portraying classroom applications, and validating what is reported. All of these activities are part of an evaluation process and communication.

Formal review mechanisms can be imposed on the development of the video to ensure that the substance of reports and communications adheres to the rigorous requirements of good evaluation. A review process can be designed to include an editorial panel, researchers as advisors, and practitioners. These people would have the responsibility of ensuring that the information presented is accurate, and that the information appropriately communicates the effects of the RTL program and how the findings benefit the educational system. Outside reviewers can be employed as impartial technical advisors and even on-screen critics or discussants. Cost controls would need to be imposed, but the expense of developing a 30 to 60 minute video documentary of studio quality could be less than the cost of developing both a conventional study with similar evaluation purposes and a video that describes the study's findings. It is likely, however, that the cost of a video documentary will vary with the overall quality of its content and imagery. The least expensive video would derive from a collection of existing video materials from RTL projects; the video and evalu-

"Formal review mechanisms can be imposed on the development of the video to ensure that the substance of reports and communications adheres to the rigorous requirements of good evaluation."

ation would be edited from these materials and presented with a sound track to communicate the range and value of RTL projects. The most expensive video would consist of original footage; the video would be of network quality and the analysis would investigate RTL's impact on classroom practices.

Research Communities Culture Analysis

An important contribution of the RTL program and other NSF programs is the development of mathematics and science education research communities. A cultural analysis could be carried out on these communities and on the links that these communities have with other relevant professional communities. The analysis of the mathematics and science education research communities could then be compared or contrasted with analyses of research communities in other subject-matter areas (such as language arts, social studies, and fine arts), other funding situations (such as the private sector or research funded by private foundations), or in other countries.

An evaluation team would be responsible for conducting the analysis. This team would—at a minimum—be composed of a mathematics educator, a science educator, and a cultural anthropologist/evaluator. In exploring the culture of researchers that has evolved through their individual interactions with the RTL program and other NSF programs, a number of questions can be addressed.

- What constitutes the research community culture that has evolved through NSF programs? Which people form the community? What is the entry into this community and how do people drop out?

- What interactions exist among the members of the community, when one considers both the mode and frequency of interactions? How do members of the community join together for cooperative work?

- What beliefs are shared by the members of the community? What support systems are in place?

- What are the patterns of migration and grouping? What are the traditions and forms of communications? Is there a common language? Are there those who would be considered outliers in the community?

- What alliances have been formed with the community and other organizations and groups? What is the power base within the community, and how powerful is the community in relation to other communities?

- What is the "gross community product" as indicated by materials produced, conference presentations, funding generated, and other measures of production?

- What are the mechanisms for transmitting the culture, and is it in the process of expanding or declining?

The main sources of information for a cultural analysis would be the researchers who have received funding through NSF and others who could be considered members of the culture (graduate students and other researchers closely aligned with members of the research community). A cultural analysis would gather information from the members of the community using interviews, questionnaires, personal resumes, and other sources used by anthropologists in studying cultures. One of the fundamental questions that would have to be addressed first in such a study concerns the actual existence of communities of researchers in science and mathematics education. Even though communities that are identified may not be considered to be "cultures" from a narrow anthropological perspective, such an analysis could produce useful descriptive information about the communities that will communicate some of the value that has been gained through the NSF programs. The methodology of cultural analysis, as used by anthropologists and others, offers the means to validate the information and conclusions that would be developed in such a study. Contrasting the research communities that have evolved out of NSF programs with other situations where other research communities have—or have not—evolved would add to the credibility of information about the importance of NSF. For example,

- One significant point of contrast might be found in the educational research communities that exist in other countries, a contrast that would be instructive despite acknowledged differences between educational systems and their relationships to local and national government.

- Another significant contrast might be found in the research communities that have formed in this country for other content areas in which no NSF funding is available.

- A third significant contrast might be found in the work and interactions of educational researchers who are supported primarily by private foundations, and in the interactions or overlaps of this group with the community of researchers funded by NSF.

- Yet another source of confirming information would be to consult the different mathematics and science education professional organizations, to ascertain the value placed by these organizations on the research communities at issue. Some indicators of this value include the visibility of the research communities in these organizations and the distribution of research findings by these organizations.

The ultimate product of the culture analysis recommended here would consist of written reports that would provide detailed profiles of research communities, their relationships with NSF, their contributions, and their uniqueness in contrast to other research communities.

Generalizability Analysis

In order to identify and examine the breadth of the RTL program's impact it would be useful to undertake a generalizability study. The purpose of a generalizability study would be to consider the impact of the program by looking at a sample of projects that have been selected randomly from those funded by the program. Although the study would reduce the costs of studying program effects by focusing on a smaller number of projects, it would have the power to suggest generalizations about the program. Certainly, the ideal situation would be to be able to study, in depth, all of the projects that have been funded by the program, and to report the effects of

"The purpose of a generalizability study would be to consider the impact of the program by looking at a sample of projects that have been selected randomly from those funded by the program."

"The four varieties of studies that have been described in this paper have been designed to provide information on a range of effects of NSF programs."

each one. However, with the nearly 200 projects funded by RTL, for example, this would be a very large and expensive task. One assumption for doing a generalizability study is that it is important to look at the effects of the program as a whole, rather than the effects of only a few projects that might be considered to have been the most productive. One reason for doing a generalizability study is that not all projects have the same scope or concentration as others. Some projects serve specific local needs; others support beginning researchers; and others may be in the very initial stages of developing an important chain of inquiry. A random sample of the projects from a program would provide a cross-section that would offer a better description of the whole program than a review of only a few, large "star" projects. How large a random sampling is needed and how the selection should be done would depend on the program and the different facets of the program to be considered.

The study of each project would require data gathering to document the effects of the projects on classroom practices, teachers, theory-building, and other applications. The expectation is that the findings from this cross-section of projects will be distributed across all of the four cells in Exhibit 1. Depending on the findings across the projects studied, statistical techniques can be used to generalize from findings common to a number of the sampled projects to all those in the program. Some supporting information on the extent of the effects of the NSF program can be obtained by using the more traditional means of administering questionnaires to a random sample of the members of targeted groups, such as the teachers' professional organizations (e.g., NCTM and NSTA), classroom teachers, scientists, and mathematicians. The purpose of these questions would be to determine what awareness members of these groups have of the NSF programs' findings, their knowledge of the findings, and the degree of implementation. This more traditional approach to evaluation is recommended in the expectation that it may determine, at some level, the range of people who are being reached by information generated by the programs. For example, a number of people are probably at least aware of some of the findings reported by Harold Stevenson from his study of Japanese, Chinese, and American students. Adherence to assumptions and conditions for doing the statistical analyses will be used to verify the findings and conclusions. The results of the generalizability analysis would be presented in a written report.

Discussion

The four varieties of studies that have been described in this paper have been designed to provide information on a range of effects of NSF programs. The four studies have been conceptualized in nontraditional ways so that they could capture aspects of the NSF programs that may be overlooked by more conventional analysis and so that they can communicate the value of the NSF programs.

- The retrograde analysis can be used to examine the effects and value of research that emerges from within a program, and to communicate a clear view of how projects within a program build on each other. If the projects within a program do not build on each other, then it is very difficult to argue that people outside of the program will be using the results. Because the research efforts of a program are directed toward developing a body of knowledge, in the absence of some internal consistency the developing body of knowledge will be fragmented.

- The video documentary approach to evaluation can very effectively communicate to a wide audience the major themes and main messages that grow out of a program. The production of a video will depend on the existence of a created body of knowledge, but it will also consider applications of work beyond the projects that fall within a program. The process of producing a coherent and precise video requires a thorough analysis of the program under investigation. Video can be a very efficient way of condensing a large amount of information in a short period of time—information that communicates the range of projects that are supported by an NSF program.

- The cultural analysis of research communities focuses on the ways that NSF programs are developing a national resource of mathematics and science education researchers. A careful explication of these communities and the operations of these communities will document and probe one of the important contributions that the National Science Foundation has made. An analysis of clearly described research communities will highlight the work of these communities in producing research and applications under NSF sponsorship; simultaneously, the analysis will report the secondary effects of experience that has been gained through work on NSF projects, and it will identify the importance of those effects to other efforts—in teacher education, writing curriculum and evaluation standards, curriculum development, assessment development, and evaluation studies.

- The generalizability analysis is designed to reveal the spectrum of effects across an NSF program by studying a sample of funded projects. This kind of study can produce information on the range of research and innovation across a program, the diverse nature of these projects, and how these projects as a collection are infiltrating the educational system both locally and nationally.

Together, the four types of evaluation study treated here would present a strong profile of the National Science Foundation to its varied audiences, and would very effectively communicate the value of the Foundation's support of research and innovation.

References

Borg, W. R., and Gall, M.D. 1983. *Educational research: An introduction.* New York: Longman.

Carpenter, T. P., and Moser, J. M. 1983. The acquisition of addition and subtraction concepts. In *The acquisition of mathematical concepts and processes*, R. Lesh and M. Landau, 7-14. New York: Academic Press.

Carroll, J. B. 1963. A model of school learning. *Teachers College Record* 64: 723- 33.

Kaplan, A. 1964. *The conduct of inquiry: Methodology for behavioral science*. Scranton, Pennsylvania: Chandler Publishing Company.

Webb, N. L., Schoen, H., and Whitehurst, S. D. 1993. *Dissemination of nine precollege mathematics instructional materials projects funded by the National Science Foundation*, 1981-91. A final report to the National Science Foundation (grant MDR 9252727). Madison, Wisconsin: Wisconsin Center for Education Research, University of Wisconsin.