The Search For Footprints: Nontraditional Approaches To Evaluating NSF's Programs

Joy A. Frechtling

The National Science Foundation (NSF) supports a number of programs that are designed to produce state-of-the-art research and innovative technical applications for mathematics and science education. Projects funded under these programs vary widely in their scope, size, and duration. Some are one-of-a-kind efforts, designed to investigate a new approach, theory, or technology. Some may be part of a stream of research, involving projects that build on each other to create a comprehensive model or those that move from theory to practice. Still others represent cooperative ventures that blend the resources of NSF with those of other funding agencies to address issues of joint interest.

While the peer review process for selection of grantees provides one important type of evaluation of NSF's programs (in the sense of quality control over what is supported), NSF, like other government and private agencies, also needs to conduct more formal program evaluations—evaluations that can be used to document the impacts and, as relevant, the shortcomings of its programs. Quality control needs to be supplemented by quality review.

However, evaluating programs such as the ones described above is neither easy nor straightforward. Traditional educational evaluation strategies that have been useful in evaluating programs that support the delivery of new services, instructional strategies, or curricula (the most familiar and widespread evaluation challenge) are not directly applicable to the majority of the research-oriented, groundbreaking inquiries that make up the portfolios of many of the Foundation's efforts. Further, the kinds of program impacts that can and should be expected of many NSF programs differ in some important ways from those typically considered where service delivery projects are the focus of study. For example,

- Traditional educational evaluations seek to attribute any impacts found to a single source, be it a support program such as Chapter 1 or a classroom intervention such as cooperative learning. For many of the programs at NSF, drawing such uni-dimensional causal statements is unlikely or impossible.

- Traditional educational evaluations have relied almost entirely on quantitative data or on counts of events. For many of the programs funded by NSF simple counts are misleading; a single successful project may justify the entire research investment, and use of quantitative indicators may exclude important areas for which no appropriate quantitative measures exist.

- Traditional educational evaluations of programs in the education sector have given priority to measures of student achievement as the impact measure of greatest concern. For many NSF programs, student achievement is an inappropriate measure either because of the nature of the research itself or the fact that any impacts on students would not be expected in the short run.

Recognizing this lack of alignment between traditional evaluation models and the nature of the programs that NSF needs to examine, the Division of Research, Evaluation and Dissemination commissioned a series of papers designed to explore alternative, nontraditional approaches to evaluation The goal is to "stretch our minds" with regard to evaluation and to explore new options, rather than to stipulate new prescriptions. This NSF project,

dubbed "Footprints," is an attempt to examine the impacts of funding programs that have been part of NSF's repertoire for a number of years and to assess the impressions they have made on the field, on scholarship, on other institutions, and on practice. This monograph presents the results of that project.

In reading the papers, it is important to keep in mind that they are not evaluations of any particular program or programs. Nor are they, in many cases, fully developed designs that could be adopted and used tomorrow or next week. Rather, they are options, speculations, and propositions that represent each thinker's ideas on how one might trace the impact of NSF's programs of support. Further, while designed with NSF programs specifically in mind, the approaches should provide food for thought for other institutions and agencies faced with similar evaluation challenges.

The papers have been solicited from a diverse group of thinkers who approach the evaluation task from both differing backgrounds and philosophies. And, because they were encouraged to think broadly in constructing their interpretations, they have produced conceptualizations of "nontraditional" that vary along a number of different dimensions.

- Some authors have emphasized the need for nontraditional evidence, indicators of program success that vary significantly from the student achievement indicators that have characterized more traditional studies. In line with this, several authors have looked for footprints in terms of effects on actual practice, on accepted models of learning, on methodologies, and even on policy.

- Some authors have stressed the development of nontraditional methodologies, supplementing the quantitative approach with one that relies more, or even exclusively, on qualitative inquiry. In fact, almost all the papers include qualitative analysis to some extent or another.

- Another dimension of difference is that of the role of the stakeholder, as opposed to the professional investigator, as the generator of hypotheses and the discoverer of impacts. Following recent trends in evaluation, almost all the papers underline the importance of stakeholder involvement—especially in understanding program goals and objectives. Some go even further and see the role of stakeholders as central to the whole evaluation enterprise.

- The papers also differ in the extent to which program evaluation is seen as an aggregation of the evaluations of different projects versus an evaluation of the program as a whole. Those who fall into the first camp seem to feel that the same outcomes that are used for assessing project success can, and should be somehow aggregated to assess program success. Others seems to follow the old saying that "the whole is greater than the sum of its parts" and seek other sources of evidence.

- Finally, the papers also differ in what could be called the "level of maturity" of the proposals being offered. Some could probably be implemented tomorrow, or at least next month, if NSF chose to do so. Others are more preliminary and will need considerably more thought and development before it is possible to assess their efficiency. These provide a core of ideas for new research on evaluation methodologies should NSF or some other agency choose to move in that direction.

Also included are a series of "reaction statements." These are not fully developed papers as such but, rather, brief statements offered in response to some of the ideas expressed. Some provide challenges to the authors; others are endorsements of an idea or point of view. The final responses attempt to put the ideas into perspective and provide suggestions for the next steps that NSF might take.

Susan Gross
National Science Foundation

Dear Reader:

The papers and discussions contained in this monograph were prepared for a conference on non-traditional evaluation methodologies that was convened by the National Science Foundation in July 1993.

NSF embarked on this project because of a need to evaluate several of its programs that were not structured in the typical service delivery model. The programs support research projects and special studies that are designed to shed light on what we know about the teaching and learning of science and mathematics. Four NSF programs were the focus of the commissioned papers:

Research in Teaching and Learning

RTL supports projects which investigate how individuals and groups learn, teach, and work effectively in complex, changing environments.

Applications of Advanced Technologies

AAT supports research, development, and proof-of-concept projects that address issues at the forefront of technology applications to learning and teaching in science and mathematics.

Studies

The Studies Program supports research projects on significant factors, trends, and practices in education, with an emphasis on their policy application.

Indicators

The Indicators Program supports studies that provide statistical information about the status of mathematics and science learning.

In my introductory remarks at the conference, I attempted to illustrate what we are looking for by use of the "footprint" metaphor. The metaphor arose from preliminary discussions concerning the four NSF programs in need of evaluation. Evaluation of these programs presented a challenge; we needed to find evidence that the programs were leaving "footprints in the sand" of mathematics and science education in the nation. Thus, the conference became known within NSF and among the authors as the "Footprints" Conference. The following ramblings are the remarks I made at the conference. The illustrations shown here were actually light-hearted computer art that was prepared for the conference—alas, they lose a bit in the translation. They are included here at the suggestion of several conference participants who felt they helped establish a focus or context for the day. I hope they work as well in print.

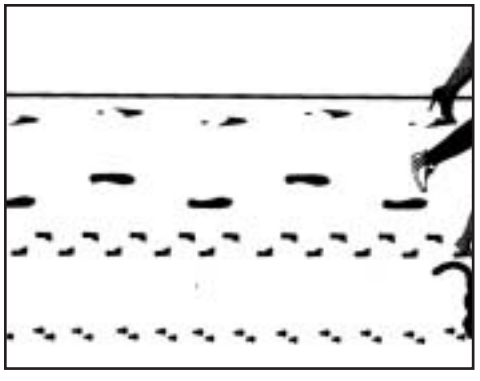
Susan Gross
Program Officer
National Science Foundation
Division of Research, Evaluation and Dissemination

Remarks

The central theme in the papers that we commissioned is Footprints.

Footprints can be viewed as the evidence of a program's impact.

Examples include: evidence that the program has had an effect on mathematics or science education; evidence that the results obtained from one or more projects funded by a program are disseminated and used elsewhere.



Footprints come in various shapes and sizes.

We should look for many types of programmatic effect, for example, changes in how we think about teaching and learning; evidence that the latest research is considered when teacher training programs are planned; examples of how the latest developments in technology are used in classroom instruction. Different types of evidence are appropriate for different types of programs. We would hope to see payoffs of research programs affecting teacher training, classroom instruction, and student learning. The production of statistical data reports, on the other hand, might result in changes in national or state policy.

Some Footprints will last a long time. This can be both good and bad.

When something worthwhile has been accomplished, it should be disseminated, replicated, and thoroughly examined and understood. However, there is the danger of a good thing hanging around too long and becoming out of date or no longer the best thinking. LOGO is an example of a computer language that served its purpose and is no longer considered state-of-the-art. Emphasis on basic skills instruction to the exclusion of higher order thinking and solving of complex problems is no longer considered the best educational approach.

The surface in which Footprints are left is important.

If the surface is not prepared adequately, the findings will be washed away. A properly prepared surface will allow lasting impressions to be made. This means that stakeholders (e.g., program planners, decision-makers, project PIs) should be involved in planning the evaluation so they will be accepting of the results.

We need to know where to look for Footprints.

How do we know we have collected all the evidence? Where are the likely places to look for missing evidence? For example: What are the untouched areas of research? What is not being done or is being done ineffectively? Are there key target groups that are not being served or are being served inadequately? What rival hypotheses can we formulate, e.g., where would we have been if this program did not exist?

We need to know when a Footprint has outlived its usefulness.

Yesterday's goal for education reform may no longer be a goal because we have moved beyond it. We need to be vigilant in retiring or making extinct old goals and adopting new ones that move us to higher standards of excellence. We must examine with regularity statistical indicators that we use to assess the health of the nation in mathematics and science education. An indicator can lose meaning because the nation has attained it, or because people work toward it as the end product rather than as the means to a larger end.