Considerations For The Evaluation Of The National Science Foundation Programs

Richard T. Hezel, Hezel Associates
Syracuse, New York

Introduction

This paper contains a set of considerations and suggestions for the evaluation of any National Science Foundation program. Of special interest to the writer is the Applications of Advanced Technologies (AAT) Program, which seeks to generate knowledge on the applications of new advanced technologies to the learning and teaching of mathematics and science. Moreover, the AAT Program strives to inform researchers, policy makers, decision makers, vendors, and developers of instructional materials about the research associated with funded projects.

An initiative that focuses on rapidly transforming technologies, the AAT Program, by its charter and mission, requires flexibility. The program accepts certain inherent risks in the funding of advanced technology projects, some of which may meet outstanding "success," while other funded projects may appear to "fail."

Program Profile

The AAT Program has supported projects whose goal is to investigate the development and use of advanced technologies, as well as projects that permit the broadest dissemination of information about the uses of technologies in various settings. AAT has supported research on, and uses of, innovative, cutting-edge technologies that have not previously been applied to particular uses in math and science education. Because the program supports advanced technologies, the program's goals, along with some of the technology applications to be supported, have changed somewhat over the years to reflect concerns with innovative, experimental technologies that might have applications in education.

By most standards of experimentation, "successful" projects yield outcomes which are desired, hypothesized, and expected. In some cases, unexpected outcomes, though not originally desired, generate results that are unforeseen, but still positive. In other cases, while the hoped-for results might not be realized, the project might yield valuable information that has long-term effects on the program and subsequent projects.

By its mission, therefore, AAT tends to support high-risk projects in which a "successful" outcome is uncertain. If successful, the projects also have the potential to provide a high payback to the education community at large. As a result, AAT has been willing to accept a higher risk and a potentially higher "failure" rate for funded projects. For the evaluator, such high risk/high gain outcomes present a challenge of assessing the value of the project outcomes, particularly when a substantial number of projects may not produce the desired results.

Part of the value of the program resides in project grantees' abilities to quickly disseminate information about their findings. Regardless of the project outcomes, the application of new technologies in learning settings requires that

"In some cases, unexpected outcomes, though not originally desired, generate results that are unforeseen, but still positive."

project results reach potential technology users as broadly and rapidly as possible.

Introduction to the Evaluation

Since 1984, NSF' s Directorate for Education and Human Resources, through its Applications of Advanced Technologies Program, has funded projects designed to generate knowledge on the applications of new advanced technologies to the learning and teaching of mathematics and science. NSF is currently engaged in planning for the evaluation of the AAT Program, and this paper has been prepared to assist that effort. Specifically, it describes potential approaches to evaluation of the program, methods that might be useful in evaluation, and special considerations for evaluation due to the innovative nature of the program. While each NSF program carries out internal evaluations, primarily through committees of visitors, this evaluation project represents the first attempt at an external evaluation of several NSF programs. As such, regardless of the frequency of evaluation, the current evaluation should be perceived as part of a system of self-renewal (Worthen and Sanders 1987), not as a discrete study oriented toward specific decision outcomes.

Program evaluations may be designed concurrently with program development or added subsequent to the program's development and initial operation. In general, the evaluator's role is more broadly defined where the evaluation is planned during the program's development. In such a case, the evaluation is collaborative with the program, administrators and grant recipients. Then, the evaluation itself is viewed as part of a continuing process in the life of the program, and all participants view the evaluation and its outcomes as central to the program's development.

By virtue of the fact that the current evaluation was conceived after the program had been operating for a considerable time, the question of what to evaluate, how to evaluate it, and what to observe presents a significant challenge. By design, the evaluation is "post hoc," in that many of the observations are made in retrospect.

Ordinarily, in the retrospective approach to evaluation, especially one that follows many years of program operation, valuable data collection and observation opportunities are lost. In particular, the opportunity to collect data that measure progress toward goals is absent. Regardless of whether the evaluation is goals-based, the post hoc evaluator has fewer options in the observation of outcomes of the program and its projects.

A concern that often arises in the evaluation process is the "intrusion" of evaluation in program design. Clearly, in the post hoc evaluation design, the evaluation cannot be said to inhibit the program design, because the program is designed independently of any evaluation plans. Therefore, despite its limitations in data collection during the progress of the project, the post hoc evaluation has substantial merit.

The post hoc evaluator has neither precedents for the design of this evaluation nor historical, systematically collected data that might contribute to it. Therefore, the evaluator is relatively free of predeterminations and biases that might have been introduced by design precedents and historical data schemata.

Approaches to the Evaluation of NSF Programs

In light of the developmental history of AAT and other programs, several alternative evaluation directions are evident. The principal approaches can be labeled broadly as objectivist and subjectivist. While a systematic, objectivist approach may be desirable for the evaluation, it may not work effectively because of the complex phenomena to be observed in such a program and in the projects that the program funds. Therefore, a subjectivist approach, which accounts for a variety of phenomena and various methods of measurement, would seem more appropriate.

To what extent should the evaluation rely on programmatic goals to set the evaluation agenda and scheme? In light of goals established for the programs, some combination of goals-based and goal-free evaluation seems warranted.

Goals-Based Evaluation

In cases where programmatic goals have been clearly established during the program's formation, the goals and the subsequent concrete and precise objectives become the criteria for measuring the "success" of the program. The goals-based approach is particularly useful for evaluating those aspects of the program that are circumscribed by goals established for the program. In this case, the goals established for the program articulate in a general way the outcomes expected from the program. In turn, the expected outcomes form the basis for the measurement of actual outcomes.

The AAT program has some general goals and objectives, which could form the basis of an evaluation. Nevertheless, a goals-based evaluation project, to be successful, requires the important intermediate step of validation of the goals as historically accurate and representative of administrators' intentions. A pre-evaluation paper summarizing the AAT goals is an important step toward a goals-based evaluation. A goals-based evaluation, in which outcomes are compared to goals, is desirable for part, but not all, of the evaluation. It is important to note that the goals-based component of the evaluation is not to be construed as utilizing a discrepancy model. The discrepancy model chiefly seeks differences or discrepancies between goals and outcomes. As a result, the model is "problem-oriented," and therefore biased toward negative evaluations.

Goal-Free Evaluation

In the absence of clearly articulated goals, or where articulated goals do not appear to circumscribe the sum of possible evaluation criteria and data to be collected, a naturalistic approach is appropriate. Such a strategy permits the collection of data from multiple sources in a retrospective manner free of the constraints of goals and their outcome expectations. Based on the description of the program and information gleaned from the clients and stakeholders, the evaluator organizes potential sources and locations of data and collects available existing and new data.

Scriven (1972) would most likely argue primarily for a goal-free evaluation, particularly where either goals are not clearly articulated or where the goals do not delineate the likely outcomes. The goal-free evaluation avoids the narrow focus of pre-established program goals and allows the evaluator to focus on actual outcomes, including unanticipated outcomes, rather than intended program outcomes only. A goal-free

"... where programmatic goals have been clearly established during the program's formation, the goals ... become the criteria for measuring the 'success' of the program."

evaluation is likely to increase the likelihood that unanticipated side effects, both positive and negative, also will be noted.

"Footprint" Evaluation

A type of goal-free evaluation that is both phenomenological and constructivist has been labeled "Footprint" evaluation. Free from the stringent limitations of traditional, management- or objectives-oriented, goals-based evaluations, the investigator examines the project outcomes not anticipated by goals. Of particular relevance in the NSF evaluation are the short- and long-term effects of the programs on their various stakeholders and nonstakeholders. The outcomes of each funded project can be observed most centrally and efficiently at the level of the project director. The broader outcomes, especially secondary influences of the project, require the evaluator to cast a wider observation net.

The assessment of dissemination efforts and outcomes especially crystallizes the trade-offs that occur in selecting either a goals-based or a goal-free evaluation approach. In favor of the goals-based evaluation, the more planned the dissemination has been, the greater the likelihood that dissemination outcomes will be traceable and identifiable. At least the evaluator has clues about where to look for evidence of dissemination attempts, so that the efforts might be assessed and future footprints will be identifiable and identified.

Dissemination Evaluation as an Example of Footprint Evaluation

The dissemination process raises other issues for the "Footprint" evaluation and provides pertinent examples for goals-based and goal-free evaluations. From the perspective of the goal-free evaluation, the evaluator observes possible dissemination outcomes, somewhat systematically and randomly, but anticipating where they are most likely to occur. The investigator searches in many and various places, not just in the places where the planned dissemination was to occur.

In particular, the effects of project information dissemination may be most effectively assessed in their potentiality, that is, the dissemination efforts attempted that are not part of the actual or real impacts of the project on the profession and the public. As demonstrated in NSF program goals, the dissemination process is vital to program success. Therefore, project dissemination attempts should compose a major portion of the evaluation.

Among the dissemination questions to be treated by the evaluation are the following:

- How and to what extent do project information and outcomes influence the variety of publics who are among the target groups of the program?
- What impact does the project have on individuals in the education profession and other institutions in terms of the development of ideas, research, and practice that emanate from the funded project?
- What new research and applications are undertaken as a direct result of the funded project and its findings?
- To what extent has the funded project yielded information that has been widely disseminated to groups and individuals in education and in business?

While the potential impact assessment of the project and its dissemination are identified above, the actual dissemination process should also be evaluated. Included in the process evaluation are the type, methods, and extent of both planned and unplanned dissemination of project results.

Evaluation Orientations

To be avoided in the evaluation of NSF programs is a utilitarian approach, which would suggest that the value of any program rises in direct relationship to the number of people the program serves successfully (House 1976). Applying such an evaluation scheme to NSF programs and their funded projects would result in the predetermination that projects that serve the most individuals, either directly or indirectly through information dissemination, would have the highest value. While the indirect influence of the programs and projects may be immense, there are limits in the ability to adequately measure the sum of the influence.

As for any program evaluation, the evaluation of NSF programs demands the recognition of one or more orientations or clients. Principal orientations or clients of the NSF evaluation are NSF administrators, consumers or taxpayers, and experts in the fields of investigation.

On behalf of the program management, the evaluator seeks to identify the decisions the administrator might make and collects useful information that demonstrates the advantages and disadvantages of each decision alternative. Program modification and improvement are examples of decision alternatives of the NSF program evaluations. The management-oriented evaluation assumes that the administrators are the clients of the evaluation and that they seek the evaluation findings.

A consumer-oriented approach to the evaluation of programs has the taxpayer-citizen as client. Through a consumer orientation, the evaluator seeks to determine the effectiveness of the program in terms of its value and service to the public, however that public is defined. When combined with a management approach and applied to NSF programs, the evaluation takes more of a public interest stance: How is the program benefitting the public citizens in general or some broad group of individuals the program is intended to serve?

Because of the esoteric nature of some NSF programs, there is considerable value in an expert-oriented evaluation, which relies primarily on the subjective professional judgment of experts in the fields of research whose outcomes are being evaluated.

The clients represented above can be considered stakeholders in the NSF programs. While a stakeholder evaluation alone, as such, is not advocated here because it lacks necessary breadth, it is important for the evaluator to consider the client/stakeholders as both sources of data and as groups to observe for the collection of data. Among the stakeholders are

1. The funders, NSF administrators, and program administrators;

2. The grant recipients and their associates who execute the projects;

3. Direct recipients of the project results or information dissemination, mostly in the academic and technology business communities;

"To be avoided in the evaluation of NSF programs is a utilitarian approach ..."

4. Indirect beneficiaries of the project results, mostly the public at large; and

5. Possible unintended "victims" of the program, such as taxpayers, groups systematically excluded from projects or their outcomes, and people who suffer negative side effects of otherwise useful projects.

Data Collection for the Evaluation

The measures of the goals-based evaluation flow directly from the operationalization of the goals, and tend to be more quantitative than qualitative. The Footprint evaluation requires a different set of data collection and methods from the goals-based evaluation. Data are collected more "naturalistically," with an emphasis on qualitative, as opposed to quantitative, data. Some of the measures, methods, and evaluation targets are described below.

Recommended Evaluation Topics and Measures

- Assess perceptions of the project, especially project outcomes, through interviews with project directors, their colleagues and associates, participants, and other experts who are familiar with the field.

- Assess the number and perceived value of new ideas and models of learning and teaching with technologies created and tested under NSF sponsorship or stimulation.

- Assess experts' perceptions of the ideas and models created and tested under NSF sponsorship.

- Assess experts' perceptions of funded projects and the value of project outcomes.

- Assess experts' retrospective and current perceived value of NSF-supported research and development on applications of advanced technologies, especially with regard to innovativeness, national impact, and uses of advance technology for learning, thinking, and problem solving.

- Assess the perceived "usefulness" and value of research on cutting-edge technology.

- Estimate the extent of uses of program-supported advanced science and mathematical concepts by educational leaders and in classrooms.

- Estimate the capacity of students to cope with problems of increasing abstraction and complexity at earlier ages.

- Analyze the results of pilot testing of new concepts and prototype materials in schools and colleges, especially with regard to the understanding of how and when new ideas can be introduced into the curriculum.

- Analyze the results of dissemination of all research completed under NSF support, including scholarly articles, articles in professional publications, news coverage, and presentation in other media.

- Undertake studies of public awareness of key concepts developed and disseminated under NSF support.

- Assess how and to what extent project information and outcomes influence the variety of publics who are among the target groups of the program.

- Assess the impact of projects on individuals in the education profession and other institutions in terms of the development of ideas, research, and practice that emanate from the funded projects.

- Determine what new research and applications are undertaken as a direct result of the funded project and its findings.

- Assess the extent to which funded projects have yielded information that has been widely disseminated to groups and individuals in education and in business. Identify the type, methods, and extent of planned dissemination of project results, and the type, methods, and extent of unplanned dissemination of project results.

- Assess the number of minority individuals participating in a program; the type, number, and effectiveness of minority outreach efforts; and the number of minority groups and individuals reached.

- Assess the program's impact on teaching and learning among individuals who have participated in the project and among individuals who have been reached by the program dissemination efforts.

- Assess among grant recipients the sources and origins of project ideas and goals, including the role of NSF funding and support in the generation of the ideas.

- Investigate follow-up activities to the grant activities, in particular what new research, projects, and dissemination have occurred.

- Track the planning of future anticipated directions and applications of the funded activities.

- Determine from principal investigators the duration of projects and the difference between the proposed and actual duration of each project.

- Assess investigators' initial goals for research and project activities, unanticipated findings that emerged from the research, and other research that has been pursued outside the scope of the grant or the project plan.

- Assess the effects of the project on participants, their attitudes, and their learning, and perceptions of the role of the project in their lives.

- Assess the impact of the projects and their activities on the professional activities of other individuals and organizations who have used the projects and their findings for other purposes.

- Conduct a thorough document analysis, including a review of each proposal and final report to determine initial goals and actual outcomes. Conduct interviews of NSF program decision makers regarding feedback received from past recipients, how past-funded project results affect future funding goals and decisions, and how the project results guide the formation of future goals.

- Assess criteria NSF uses to determine the "success" of projects and how NSF decision makers arrive at the criteria.

- Assess the methods NSF programs use to decide which projects are to be funded. Determine what predictors of success are applied from past projects.

Conclusion

The paper has offered recommendations for the evaluation of NSF programs. Given the posthoc nature of the evaluation design and the absence of identified measurement criteria, an evaluation that is solely goals-based carries serious limitations and is ruled out. Instead, a combination of evaluation approaches that includes both goals-based and goal-free methods is necessary and recommended. In general, the goals-based approach is seen to be valuable in the measurement of anticipated project outcomes, while the goal-free approach assesses broad effect, including unanticipated effects. Both approaches utilize quantitative and qualitative data.

References and Citations

Guba, E.B., and Lincoln, Y.S. 1981. *Effective evaluation*. San Francisco: Jossey-Bass.

House, ER. 1976. Justice in evaluation. Vol. 1. In *Evaluation studies review annual*, ed. G. V. Glass. Beverly Hills, CA: Sage.

House, E.R. 1983. Assumptions underlying evaluation models. In *Evaluation models: Viewpoints on educational and human services evaluation*, eds. G.F. Madaus, M. Scriven, and D.L. Stufflebeam. Boston, Mass: Kluwer-Nijhoff.

Scriven, M. 1972. Pros and cons about goal-free evaluation. *Evaluation Comment* 3(4): 1-7.

Talmage, H. 1982. Evaluation of Programs. In *Encyclopedia of educational research*, 5th ed. H.E. Mitzel, New York: The Free Press.

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