Footprints On Surfaces: A Nontraditional Approach To Evaluation Of National Science Foundation Programs

M. Christine Dwyer, RMC
Research Corporation

Introduction

This paper explores an approach for nontraditional evaluation of National Science Foundation (NSF) programs that deals directly with the impact of those programs on selected organizations engaged in education reform. The proposed approach advocates examination of a "slice" of the larger picture of educational change, focusing on selected stages and actors along the continuum from knowledge development to dissemination through implementation and reform. The examination would yield information about the stage linking the knowledge generated by NSF programs to implementation. The process would trace the influences and uses of that knowledge by intermediary organizations that have training and technical assistance functions, such as teacher training institutions, educational laboratories, and state departments of education. The basic idea of tracing influences on intermediary organizations is carried through in evaluation questions, variables, criteria for selecting a sample, and data collection processes. The paper illustrates the viability of the plan through an extended example and suggests some ways to address methodological problems.

This evaluation idea fits best with the purposes of those NSF programs designed to generate knowledge about the teaching and learning of mathematics and science to inform the work of researchers, policymakers, developers, and teachers. Several characteristics of NSF programs have inspired the design, including the following:

- The goals of creating a base of knowledge applicable to learners at all levels and useful to education reformers;
- The value placed on direct utility of projects for education;
- The targeting of underrepresented groups;
- The concern for systemic change;
- The variety of projects funded and the resulting array of outcomes and types of knowledge generated;
- The high profile among practitioners of many projects and their personnel;
- The collaborative nature of funded projects, which suggests multiple paths of project influence; and
- The emphasis on innovations.

Those characteristics also suggest the major challenges for evaluation design: the difficulty of capturing important, systemwide influences; the need for a new set of assumptions to replace traditional attribution concepts; the elusiveness of effects; and the need to separate development and dissemination for evaluation purposes. A study of the effectiveness of dissemination is not intended here. Lessons from the study of policy and program implementation over the past 20 years, along with our own

"The proposed approach advocates examination of a 'slice' of the larger picture ... focusing on selected stages and actors along the continuum from knowledge development to dissemination ..."

"The purpose of the evaluation is to learn more about the varied paths and processes by which NSF programs influence educational practice, through a look at the impact on particular intermediary organizations ..."

experiences, have taught us that basing this evaluation on the programs' direct impact on educational practice would not be fair. So while this paper looks for connections to education practice, it is not intended and should not be interpreted as an evaluation of the dissemination or implementation of NSF projects.

In the next section, the evaluation purpose is discussed, with the goals of being fair to original NSF program intentions and also useful to policymakers. The section also includes an overview of the approach with special attention to explaining the concept of intermediaries. Following that is a summary of background influences that shaped the approach: the logical extensions of the Footprint metaphor; some applicable lessons from research about the influence of knowledge on policy and practice; and the author's experiences with the operations of technical assistance intermediaries. A framework for an evaluation plan along with sample evaluation questions and a discussion of the nature of study results, follows. Finally, an extended example is presented, and practical issues to be encountered in carrying out the evaluation are discussed.

Key Features: Purpose and Rationale Overview, and Role of Intermediaries

The purpose of the evaluation is to learn more about the varied paths and processes by which NSF programs influence educational practice, through a look at the impact on particular intermediary organizations that have the mission of linking research and practice for reform. The evaluation examines how the knowledge generated by NSF programs has affected or been incorporated by selected intermediaries within the larger education system. It focuses on those intermediary organizations with missions connected to systemic reform of mathematics and science teaching and leaming. Simply, stated, if knowledge was originally generated for the purpose of such reform, the question is how and to what extent active reformers have acquired and used the knowledge.

The proposed evaluation emerges from a "macro"-level perspective of how knowledge[1] changes practice, yet focuses on one element of the system of influences surrounding the knowledge generated by NSF programs. Instead of looking directly at effects on practice at the classroom or institutional level, it examines the effects on the larger system that supports, influences, and changes the work of education practitioners.

Intermediaries are agencies such as technical assistance centers, universities, teacher institutes, and laboratories with established dissemination, training, and reform functions. They serve both linking and leadership roles and bridge the cultures of research and development and educational practice through materials development, training, and networking. They are proactive in seeking knowledge generated by the research and development community. Intermediaries include the educational laboratories, the content-related Office of Educational Research and Improvement (OERI) research centers, technical assistance centers with categorical reform missions such as the 16 Chapter 1 Technical Assistance Centers (TACs), state departments of education, Federally supported project dissemination networks such as the National Diffusion Network (NDN), selected Statewide Systemic Initiatives (SSI), state and university projects for teacher training supported by the Eisenhower Mathematics and Science Education Program, universities that prepare teachers, and professional associations Of greatest interest for this paper are those organizations with the closest con-

[1]The term "knowledge" used throughout the paper is shorthand for the object of the evaluation—the myriad outcomes of project work, the ideas, principles, strategies, concepts, papers, curriculum manuals, software, materials, research results, etc., that form the work of the NSF programs.

nections to the reform of mathematics and science education.

Evaluation Overview- The proposed evaluation would (a) illuminate the paths and processes by which knowledge generated by NSF programs is selected, acquired by, and transferred to intermediaries; (b) describe the knowledge that is of interest, and not of interest to intermediaries; and (c) learn what functions that knowledge has served for intermediaries. Other, possible evaluation purposes deal with the processes used by intermediaries to translate and transform knowledge and then the experiences of intermediaries in influencing education practitioners. The sample evaluation questions below suggest what could be learned from brief case histories of both intermediaries and the paths of influence of particular knowledge examples:

- How have regional Chapter 1 TACs used NSF-supported work in the teaching of elementary mathematics to improve programs serving disadvantaged students? Do the materials used by TACs include the principles and practices that emerged from the work on cognitively guided instruction, for example?

- To what extent do any techniques developed by specific NSF programs appear in the programs promoted and funded by the Department of Education's National Diffusion Network?

- Has Eisenhower-supported state-level teacher inservice been shaped by the knowledge generated by NSF programs?

The questions suggest the components of a model framework (i.e., objects, respondents, data collection processes) to bound data collection. Clearly, the evaluation process requires heavy involvement of at least some grantees and NSF in defining the information to be tracked, hypothesizing the varied influences of particular work on practice, and identifying the intermediaries that would be both likely and unlikely candidates for influence. Therefore, a component of the approach includes work with grantees to identify the presumed paths of influence of their work. The cluster evaluation method for identifying common outcomes would be relevant (Barley and Jenness, 1993) for identifying common paths of influence. The proposed data collection processes are akin to investigative journalism approaches (Smith, 1981; Cuba, 1981), tracing leads about whether people in intermediate agencies are familiar or unfamiliar with, have used or not used, knowledge generated by NSF programs.

It is easy to anticipate arguments about this approach One could argue that because the explicit intentions of NSF grantmaking did not (and should not) include the expectation of leaving traceable marks on practice, it is simply not valid to look for effects later. Or, from an instrumentalist perspective, one might assume that, because the intentions of knowledge developers may not have been specific uses of knowledge, it will simply be impossible to trace the processes by which that knowledge was acquired and transferred within the larger system. Finally, the anticipated elusiveness of information as a result of interpretation and translation over time may make the approach seem overwhelmingly complex to some.

On the other hand, it is very easy to imagine that policy makers and decision makers at all levels might expect an eval-

"There are several compelling arguments for using technical assistance and reform intermediary agencies as the 'surface' on which to look for footprints of influences from NSF programs."

uation to answer the question of what and how and how much NSF programs have contributed to improved educational practices. The current climate of educational reform spurs everyone's interest in the extent to which changes in practice have occurred. The clear and widely promoted statements of needs for reform at all educational levels in mathematics and science teaching/learning and expansion of use of technologies have created a context in which there will be increasing pressures to look diligently for the mark of NSF programs directly on educational practice—and beyond, at student and societal outcomes. Further, because of the scope and depth of the current concern about reform, one can anticipate pressures to look for those footprints on the "biggest surface" possible, perhaps even a national landscape—hence, the interest here in considering the larger systems that support the influence of knowledge on practice.

More about Intermediaries- As is clear by now, intermediaries are a critical element of the evaluation design. Obviously, the value of using the approach depends partly on how possible it is to achieve agreement about which intermediaries act as primary channels or paths linking research and educational practice. Will grantees and NSF agree that it is both fair and valuable to trace and describe effects on the functions, understanding, beliefs, and attitudes of intermediaries? How complex will it be to attain agreement on which intermediaries are appropriate? While responses to those challenges are unknown to us at this point, it is a relatively simple matter to gather initial reactions. There are several compelling arguments for using technical assistance and reform intermediary agencies as the "surface" on which to look for footprints of influences from NSF-sponsored programs.

- First, from some perspectives, intermediaries represent manipulatable levers of change; in the spirit of systemic reform, it is critical to know how and to what degree they are influenced by and take advantage of the knowledge generated by NSF programs. They are likely to recognize and discuss the influences on their work, if any, of selected NSF programs because their espoused missions are to influence practice (whether by training, consulting, or product development) and to do so, they must be proactive in seeking knowledge and research.

- Second, intermediate agencies offer a potential solution to the problem of tracing isolated and discrete effects on practice and/or entirely avoiding looking at the effects on practice because of the complexity of where to look. Because of their multiple functions, intermediate agencies are likely to have had varied opportunities for contact with the knowledge generated by several NSF projects. For example, a regional educational laboratory initiative may have incorporated specific examples of technology use, as well as assessment practices and curriculum examples in its work with teachers.

- Third, because intermediaries are in the business of transforming research into materials, training, experiences, policies, and expertise for the purpose of influencing educational practice, they will be able to offer a rich perspective on the process of acquiring and using knowledge, and describing their own paths of contact and develop-

ment, including how they have come to know and value NSF program material. Well-selected intermediaries would be expert reporters on the entire system of influences that connects knowledge generated by NSF programs to educational improvements.

- Fourth, depending on the intermediary, there may even be some limited opportunities to estimate the effects on the broader field of practice through internally maintained client databases. A hypothetical example would be finding out the number of teachers trained in a particular set of teaching techniques developed by NSF programs and incorporated in National Diffusion Network physics programs. This is a simple matter in the case of the NDN, because information about teachers trained by specific programs is a data element maintained in a central database.

- Finally, agencies with technical assistance functions are of special interest because they generally assume a proactive role that increases the likelihood of contact with NSF-generated knowledge. That proactivity is manifested in the "scanning" associated with technical assistance agencies; by design, they are searching continuously for emerging issues and perspectives within a number of environments. Further, technical assistance interests draw upon the varied worlds of research, policy making, and education practice. Thus, technical assistance intermediaries are likely to find useful a wider variety of types of knowledge generated by NSF programs than other agencies that may be interested exclusively in direct use training materials or research to shape policy.

Developing a Perspective: Influences on the Approach

The Footprint Metaphor- The metaphor of the footprint is a helpful starting point for thinking about reasonable boundaries for an evaluation, the nature of evaluation questions, and some options for data collection. "Footprint" signifies a mark or effect that will remain visible, at least for a certain time period. The footprint metaphor also suggests an evaluation that is concerned about what marks are made, how marks are made, and whore they can or should be found. The metaphor suggests that the impressions left by an NSF program may be of varied depths, more or less visible, and more or less lasting. Much of the variation in impressions has to do with the other part of the metaphor: the surfaces on which the footprints fall. The approach in this paper emphasizes looking for the most appropriate (and one might argue, the most important) surfaces among the candidate intermediaries, meaning those that are most likely to accept, hold, and then even preserve footprints. The surfaces proposed here are examples from the national, regional, and state agencies or interest groups that have educational dissemination and reform support functions. In Karen Seashore Louis' (1981) terms, they are agencies that have external agent functions and multiple roles related to knowledge utilization: decision making, enlightenment, and capacity building. The enlightenment function (Weiss, 1972) of providing information and using research and development knowledge is especially relevant to the roles played by intermediaries as links in the research into practice continuum and to the type of knowledge generated by NSF programs. Technical assistance missions

suggest that the relationship between the intermediaries and educational practitioners is ongoing, characterized by gradual infusion of improved information and gradual learning and change.

Relevant Lessons from Research- The lessons from knowledge utilization and implementation research that may be most pertinent to this evaluation are cautions about what not to expect for outcomes, heightening sensitivities about what would be of value and interest, where to look, and what to expect. First, it is useful to review a few lessons from Milbray McLaughlin's influential summary (1987) of two decades of implementation research:

- We know to expect enormous variations in how knowledge is used, even when the object at hand is as bounded and prescribed as a packaged curriculum;
- We know that local capacity, motivation, and beliefs are. the central influences on what gets implemented; and
- We know that it is individuals within organizations who use information, reflect on attitudes, and implement changes (not the organizations as units).

These lessons suggest modest expectations for knowledge use by intermediaries and practitioners. At the same time, McLaughlin's lessons suggest we need to ask what kinds of choices, interpretations, and transformations are made to meet the information needs of different actors at different points in time. They offer intriguing possibilities for questions about the capacities, motivations, and decisions that face intermediaries as they select and shape knowledge to influence practitioners. McLaughlin's "implementing system" notion suggests attention to the connection between the knowledge generated by NSF programs and those most likely to seek and make important use of it.

Research about the utilization of social science information offers other relevant lessons to frame the questions to be answered by an evaluation:

- Since utilization of knowledge takes many forms, and is seldom used in direct instrumental ways, the relevant questions related to use are when, under what circumstances, and how (Nelson, 1987).
- When viewed from a communications perspective, the important variables related to use are source, message, channel, (the path and form of information), characteristics of the receiver, and conceptual impact (as opposed to instrumental) (Nelson, 1987).
- Utilization value depends partly on strategic conditions—timing, feasibility, values, and power orientation (van de Vall, 1987).

"The knowledge utilization literature also points to the importance of the characteristics of what knowledge gets used and the conditions surrounding use."

The knowledge utilization literature also points to the importance of the characteristics of what knowledge gets used and the conditions surrounding use. The variety of conditions surrounding the paths of knowledge use traceable to intermediaries is great. The ideal result from this type of evaluation is a deeper understanding from selected cases of how knowledge comes to be valued and used by intermediaries.

Context: The Author's Perspective- The design choices proposed in this paper about what would be both interesting and important to evaluate have been strongly influenced by my own work as a

technical assistor in national educational dissemination and reform efforts.

In large part, my work and that of my RMC Research colleagues has been about support for reform of practice in teaching and learning at state and local levels, usually functioning as the type of dissemination/reform intermediary described in this paper. Our work for Federal and state governments and foundations is about the promotion of research-based policies and practices through training, consultation, and product development. As a group, we serve in several capacities as intermediaries, translating research into practice and supporting or facilitating educational improvements that contribute to systemic reforms. These responsibilities have directly impressed upon me a respect for the challenges involved in "leaving a mark" of any type on practice—even when one understands the complexities involved in influencing changes in educators' behaviors and attitudes, and is immersed in the policies and procedures of school systems.

At the same time, it is also clear that desired reforms do occur in some situations and under certain circumstances. And it is also clear that intermediaries have played a variety of roles in the reform process: stimulating dialog, providing background information, planning evaluations, interpreting results, working in partnership with schools to identify and implement changes, creating experiences to force disequilibrium, training teachers, linking with model programs, etc. My own experience has raised interest in (a) the proactive roles that the actors within intermediate agencies play in the transformation and transfer of knowledge into practice for the purposes of reform, especially now that the reform talk has turned systemic; (b) the processes by which we intermediaries shape, renew, and revamp our own knowledge bases; and (c) the group and individual decision making within intermediaries for selecting and sharing knowledge with practitioners.

Conceptual Framework: The Questions Addressed by the Evaluation

The conceptual framework for the evaluation design begins with a "macro" view of how knowledge affects practice. Exhibit 1 is a preliminary conceptual framework, illustrating components of a model with the following characteristics:

- Within the array of NSF grantees, specific elements of program-generated knowledge will need to be selected and described for tracking purposes;
- The paths of knowledge acquisition and transfer can be simple or multiple, circuitous or direct, connected or unconnected, curious, unpredictable, serendipitous, mutual—and are best traced through exploratory, investigative activity; the path-arrows on the diagram are meant to illustrate the wide variety of patterns that might be found;
- Intermediaries vary in scope, importance, function, and role;
- Intermediaries seek knowledge from and are influenced by many sources, including NSF programs;
- Intermediaries use a variety of modes to influence educational practice; and

- NSF programs and the intermediaries are depicted within the field of education practice; obviously, both are also influenced by other elements of the field (although this is not depicted in the diagram simply to keep the discussion simpler).

As with any diagram of this type, this framework risks making to relationships among components seem less complex than they really are, but it does help to generate evaluation questions.

Evaluation Questions- At the simplest level, the basic evaluation question is about the very existence of footprints: Do knowledge footprints associated with NSF programs appear when the work and operations of intermediaries are examined? While practically challenging, whether the marks are found or not, this question is unlikely to yield information that is helpful to the ultimate purpose of NSF programs; that is, building a knowledge base that contributes importantly to reform of educational practice. Rather, the most interesting and useful questions involve asking where the footprints appear and about how they got there; what varied paths the footprints have taken; and the shape, size, and depth of the marks when located. These questions are important because dissemination paths were not originally prescribed, and therefore grantees' intentions about knowledge use are likely to vary dramatically. The utility of this evaluation is learning how knowledge reaches intermediaries; how intermediaries understand, select, and transform that knowledge; and how and to whom intermediaries promote the results. The special feature of the proposed approach is tracing both forward and backward; that is, following the paths of influences both in those cases where NSF program grantees intended dissemination for particular uses and in those where no proactive dissemination was intended.

Exhibit 1 lists four broad evaluation questions, corresponding to the relationships and processes depicted. The first two questions (What is the nature of knowledge that reaches intermediaries? and What are the paths and processes of acquisition and transfer?) seem most relevant. The question of how intermediaries translate and shape knowledge occurs at a different stage of the system of influence and is probably beyond the scope of this evaluation. The fourth question (How is knowledge used by intermediaries?) should be addressed to the extent of learning about intermediaries' intended uses of knowledge and their proposed strategies for influencing use. A beginning list of variables associated with the three questions (I, II and IV) of primary interest follows. Obviously, it would be important to involve stakeholders in identification and selection of the variables.

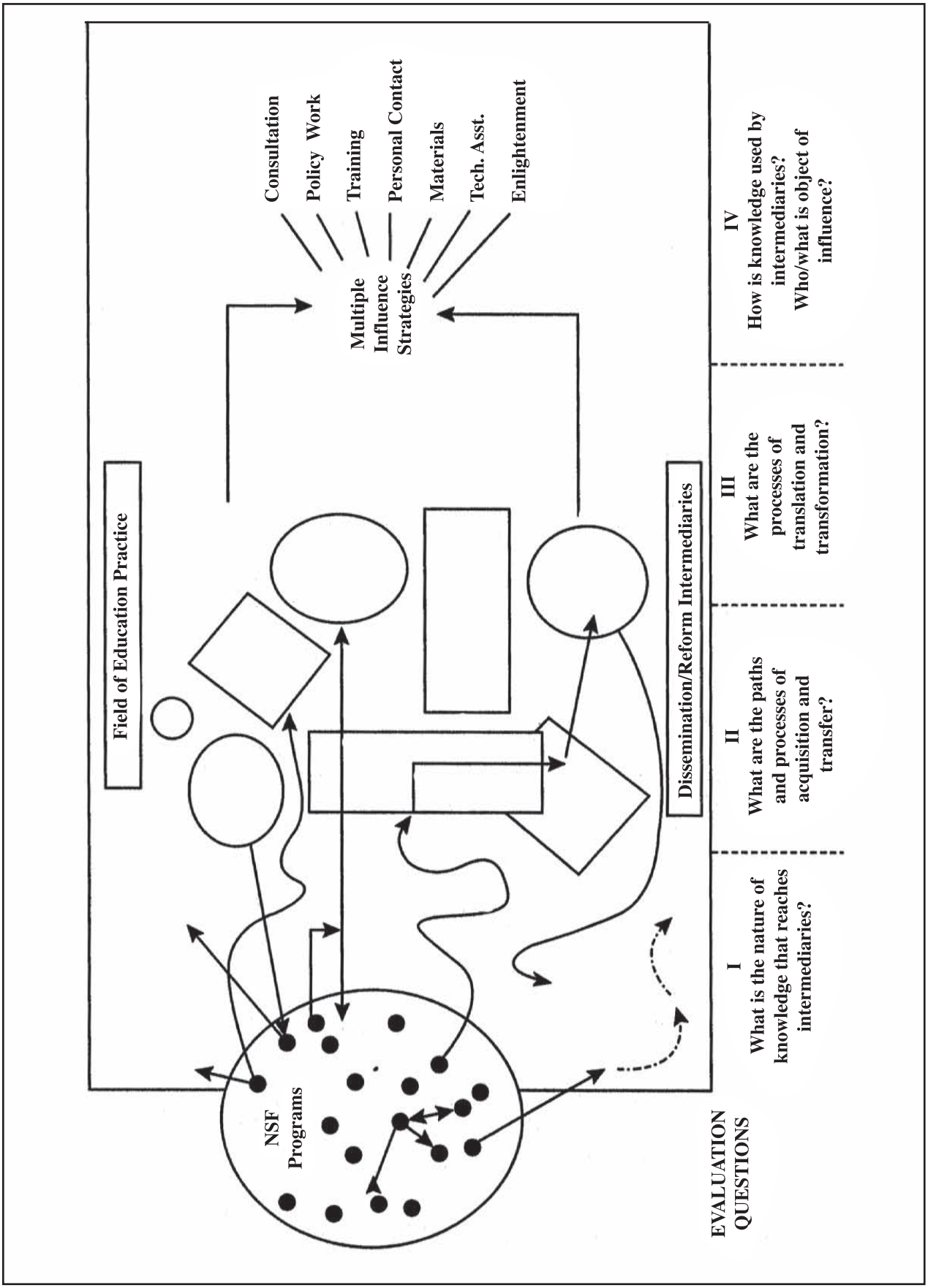

Exhibit 1. Conceptual Framework For Evaluation and Abbreviated Evaluation Questions

I. What is the nature of NSF program-generated knowledge that reaches intermediaries? What is the nature of the knowledge that does not? What are the differences?

Sample variables:

- Scope of implied change/impact
- Perceived proximity to typical practice
- Perceived and actual technical difficulty of application
- Perceived and actual implementation difficulties
- Perceived and actual degree of innovativeness
- Length of time available
- Producing institution and its affiliations
- Content
- Level
- Form, i.e., degree of "packaging" for practice
- Variety of channels and opportunities through which knowledge is available
- Amount of time investment required for initial understanding
- Directness of connection to national/state policies
- Directness of connection to student outcomes

II. What paths and processes do intermediaries use to acquire and receive NSF program-generated knowledge?

Sample variables

- Motivations and purposes for transfer
- Motivations and purposes for acquisition
- Direction of initiation
- Characteristics of initiators
- Roles and positions of key actors
- Formal relationships that facilitate transfer
- Forums for awareness and exchange
- Roles of professional associations
- Roles of colleges and universities
- Differences in initial and subsequent contacts with knowledge
- Similarity/difference with other acquisition activities of intermediaries, especially those related to mathematics, science, technology
- Barriers (attitudes, understanding) from multiple perspectives
- Role and context of personal contact
- Function of receiving unit within intermediary
- Perceived satisfaction
- Content expertise of receiver

IV. How is NSF program-generated knowledge used by intermediaries? What are the intended uses and strategies that connect to education practice?

Sample variables:

- Internal and external enlightenment functions

- Basic modes/strategies of influence
- Similarity of mode to typical strategies
- Placement within ongoing functions
- Fit within ongoing conceptual work
- Fit within ongoing instrumental work
- Stimulus for new approach/new activity/revamping
- Facilitation of connection to different levels of practitioners
- Perceived satisfaction
- Additional needs associated with intended uses

Study Method- Case histories are a logical data collection strategy, based on multimethod exploratory investigations that trace (a) forward from knowledge examples provided in NSF projects and (b) backward from selected intermediaries. Several case histories would yield a detailed picture of some of the effects of knowledge on intermediaries, and by extension, the effects on education practice. Similarly, the study would also identify knowledge that did not reach intermediaries. Judgments about the value of the emergent patterns of knowledge use and non-use become a stakeholder problem, but one that might be resolved through other parts of the evaluation—perhaps, for example, the independent expert assessments of the value of NSF project work that were suggested by several other paper authors. Cross-case analyses (using the intermediary as the unit of analysis) would provide information about how knowledge is or is not acquired and transferred.

An Extended Example and Some Practical Problems

We have not yet addressed the scale of the evaluation. A modest but indepth exploration of three to four well-selected intermediaries would be sufficient to (a) learn about the value of the approach and (b) gather enough leads about influences on intermediaries to preview effects. Obviously, the selection of intermediaries is critical; consensus on their representativeness, potential for depth and breadth of contact with NSF program-generated knowledge, and effectiveness in technical assistance and reform must be established among stakeholders early in the evaluation. The intermediaries should probably represent a wide range in terms of likelihood of use of NSF program-generated knowledge, ranging from obvious choices (i.e., those with direct and primary roles in the reform of mathematics and science education practice and the application of technologies) to those with strong and important connections to practice but less obvious connections to NSF programs.

The extended example that follows is an unlikely intermediary, chosen to illustrate the potential of the approach to uncover effects. The example previews the issues that will arise in identifying and selecting candidates and collecting data. The sample intermediary is the national network of Chapter 1 Technical Assistance Centers (TACs) and Rural Technical Assistance Centers (R-TACs). Its selection was based on the author's experience with TAC operations and not because it necessarily represents an optimal candidate. TACs are unlikely intermediaries because they are not charged

"... The study would also identify knowledge that did not reach intermediaries."

with reform of mathematics and science teaching.

The TAC network comprises 16 Federally supported multipurpose centers (approximately 65-70 full time equivalent) serving state and local education agencies in the areas of Chapter 1 program design and improvement and program assessment. In the past 5 years, TAC activities at the local level have focused largely on improvements in Chapter 1 programs, including promotion of research-based strategies for teaching and learning in mathematics, reading, and writing. The ultimate beneficiaries of Chapter 1, and therefore TAC activity, are disadvantaged students and their parents at all levels. In elementary mathematics, for example, TAC activities might well include any or all of the following:

- Identifying curriculum and materials;
- Providing awareness of the National Council of Teachers of Mathematics (NCTM) standards directly to practitioners;
- Consulting with districts to establish staff development structures;
- Training administrators, teachers, and curriculum specialists in research-based principles, strategies, and techniques;
- Conducting demonstration lessons as part of inservice work;
- Helping to locate or design alternative assessments for problem solving;
- Researching the practices of other states regarding criteria associated with standards;
- Introducing parent leaders to the principles associated with advanced skills in mathematics, and defining high expectations by example;
- Developing research syntheses to inform policy development at the state level;
- Representing compensatory education interests on a statewide committee interested in reform;
- Encouraging a district to address the weaknesses of mathematics curriculum when developing a program improvement plan;
- Helping districts interpret the implications for instruction of the results of mathematics assessments;
- Writing a newsletter article on high-powered strategies for disadvantaged learners;
- Gathering information about user experiences with particular software; and
- Developing an agenda and locating presenters for a regional or national conference on mathematics teaching strategies for disadvantaged learners.

Certainly, TACs are not the only resource that Chapter 1 programs turn to for support in reform of the teaching of mathematics. However, because the TACs are multipurpose, credible, and provide services at no cost, there is a tendency for Chapter 1 clients to contact them for a wide variety of functions, as the above list demonstrates. As a result, TACs connect knowledge and research to practice in mathematics to an extent far greater than a

"TACs connect knowledge and research to practice in mathematics to an extent far greater than a passing acquaintance with the TAC network might suggest."

passing acquaintance with, the TAC network might suggest. Furthermore, they are working with practitioners who serve an especially important group of students—students who are economically and educationally disadvantaged.

TACs have several other features that raise issues about what makes a good candidate to be an intermediary:

- Only a small proportion of TAC staff (perhaps 10 percent) have academic backgrounds related directly to elementary mathematics, so the need for acquisition of knowledge to serve clients is clear;
- Materials and training are shared across the network through established mechanisms (quarterly meetings, institutes and seminars, an electronic network, materials clearinghouses, some common policies related to materials development, a culture that supports exchange) so influences can spread fairly rapidly; and
- Two separately funded support centers for the TACs have the mission of acquiring knowledge related to curriculum and instruction and organizing, translating, transforming, and disseminating it for use by all the TACS.

The question of whether or not the TAC network would be a viable candidate for intermediary status in this evaluation can probably be answered by the degree to which the reader is intrigued at this point to find out how TACs have been using the knowledge generated by NSF programs. Intermediary selection criteria emerge from consideration of advantages and disadvantages of the TACs as candidates for this study.

The advantages are the national scope of TAC influence; the mission of improving educational programs for the disadvantaged at all levels; the multiple functions of training, policy support, planning, consultation, and product design; the simultaneous work at different levels of educational practice (classroom, school, program, district, state, regional, and national); some degree of commitment for generating improved knowledge for practitioners; the system support for enlightenment and capacity-building uses of knowledge; the relatively small size of the network and accessibility of personnel; and the capability of tracking activities through content-based client service records. The disadvantages are the multipurpose TAC mission; competing obligations, because the TAC agendas are determined largely at state and Federal levels; and the variability of knowledge use across TAC centers (as a result of organizational context and cultures as well as regional needs and interests). Those advantages and disadvantages offer a preview of criteria that might be used to select intermediaries for study.

Speculating on the results of an exploratory review of NSF influences on TACs leads to these hypotheses: the influences would be numerous; TACs would probably be the initiators of knowledge acquisition, using some traditional awareness vehicles but often becoming aware of specific knowledge through policy-related channels (Federal policy studies, for example); the primary intention of knowledge use by TACs would be teacher training through their influence on program design and policy development; knowledge with the clearest connection to student outcomes would be preferred; TACs would expect to translate research findings into best practices before using them with practi-

tioners, even for enlightenment functions; and TACs would have a strong interest in assessments and perhaps initial contact would have been based on interest in assessment.

Selecting Intermediaries- The extended TAC example of an unlikely intermediary candidate raises interest in thinking about intermediaries that would be viable. The example also introduces a host of practical problems to be faced in the study, beginning with the process of identifying and selecting intermediaries. Different stakeholders will have preferences for different types of intermediaries. The essential criteria for intermediaries should be potential and credibility—potential in terms of likelihood of locating effects and credibility in the sense of scope and importance of influence. Other related criteria are national profile and scope of influence; longevity and stability of the intermediary; clarity of mission with respect to dissemination and reform; proactivity of outreach and extent of collaboration with other intermediaries; multiple functions, including a research and development capacity; multiple entry points from the perspective of practitioners; and maintenance of records that permit tracking of client contact at some level.

"... The evaluation process is more like an investigative dialog with intermediaries than a survey of use ..."

A related issue will be identifying the best informants or reporters from each intermediary, recognizing that the functions of knowledge acquisition, transformation, and use are probably carried out by different units within an intermediary.

Bounding Data Collection- This is perhaps the most elusive element of the proposed approach. There is little guidance for knowing what program-generated knowledge would be best for tracking purposes. Selecting and defining knowledge would involve those individuals or groups most familiar with the knowledge generated by NSF projects, especially the grantees themselves. Grantees are in the best position to know the aspects of their work that have potential for influencing practice and to identify what has already found a way into practice. We envision a process that engages grantees and other stakeholders (e.g., NSF, selected intermediaries) in developing a set of theories about the presumed paths of influence associated with knowledge generated from their work. That process would yield a range of types of knowledge to be developed into descriptions for tracking purposes.

Because a key evaluation purpose is to learn primarily about the process of acquiring, using, and valuing knowledge, it would be important to select examples of knowledge that are concrete, as well as examples that would be more difficult to track. Ideally, the pool of descriptions would vary at the outset and from 0 perspective by format, scope, content and level, proximity to practice, longevity, perceived innovativeness, and technical complexity. Because the evaluation process is more like an investigative dialog with intermediaries than a survey of use, descriptions need only serve as conversation starters, not complete catalogs of program-generated knowledge. An obvious challenge is that intermediaries will have translated and transformed the knowledge as they have incorporated it into their work.

Data Collection Procedures and Analysis- The exploratory nature of tracking the influence of knowledge suggests use of the investigative journalism metaphors and models described by Smith (1981) and Guba (1981). In Guba's terms, the goal of tracking the paths of influence is to develop "working hypotheses embedded in thick descrip-

tions." Evaluators follow the trail of a chain of events, continuously using creative strategies to develop new sources and leads. The process requires the establishment of a record, reconstructing and then verifying the tracks. Next steps always proceed from what has been previously documented, analyzed, and summarized. Continual recycling to previous sources and leads with newly generated hypotheses is part of the process, as is running information back through contacts for confirmation or refutation. Data collection includes records review and analysis, key interviews, and observations to "establish a record" of transactions, profiles, chronologies, and relationships. Developing and refining hypotheses about what and how influence occurs is a matter of cross-referencing the varied pieces of information in the rich database built from the experiences of the intermediaries and the points of contact with NSF projects. As data are collected, the conceptual framework would be refined through reflection on the evolving hypotheses. Finally, cross-case analysis (each intermediary is a case) would be based on a revised framework. Both cross-case results and descriptions of the experiences of each intermediary represent valuable products.

Alternatively, one might organize and vary data collection by stages, beginning with surveys and/or focus groups of grantees to learn first about possible intermediaries, presumed paths of influence, and dissemination intentions.

Next, paths might be traced backwards from intermediaries as discussed above. Grantees with especially clear, deep, and varied connections to knowledge could then be studied in depth in another set of cases.

Summary- Suggestions provided by authors of other papers in the series offer potential solutions for methodological issues that arise from this evaluation approach. Cluster evaluation techniques could be used to define knowledge and hypothesize paths to trace. Expert judgments are required at several points to give value to the findings about paths and processes. Generalizability analysis as discussed by Webb offers some ideas for sampling the breadth of NSF programs. Several elements of Yin's partial comparison model provide conceptual guidance for thinking about the legitimacy of approaches use of proximal outcomes where interventions are weak, direct assessment of process logic, and the value of compelling explanations of documentable chains of events.

The connections to other papers reinforce the possibilities for using this evaluation approach in conjunction with others. The approach is offered as one of several "small wins" in Karl Weick's terms—an option for breaking down the complex nontraditional evaluation problem into a series of achievable tasks.

References

Barley, Z.A., and Jenness, M. 1993. *Conceptual underpinnings for program evaluation of major public importance: Collaborative stakeholder involvement.*

Bickman, L., ed. 1987. *Using program theory in evaluation*. San Francisco: Jossey-Bass.

Bryk, A.S., and Hermanson, K.L. 1993, Educational indicator systems: Observations on their structure, interpretation, and use. In *Review of Research in Education*, vol. 19. Washington, D.C.: AERA.

Guba, E.G. 1981. Investigative journalism. In *New techniques for evaluation*, N.L. Smith, (ed.). Beverly Hills, CA: Sage Publications.

Knapp, M.S., Shields, P.M., and Turnbull, BJ. 1992. *Academic challenge for the children of poverty*. Washington, D.C.: U.S. Department of Education.

Knapp, M.S., and Turnbull, B.J. 1991. *Better schooling for the children of poverty: Alternatives to conventional wisdom*. Berkeley, CA: McCutchan.

Knapp, M.S., Zucker, A.A., Adelman, N.E., and St. John, M. 1991. *The Eisenhower Mathematics and Science Education Program: An enabling resource for reform*. Washington, D.C.: U.S. Department of Education.

Little, J.W. (Forthcoming). Teacher's professional development in a climate of educational reform. *Educational Evaluation and Policy Analysis*.

Louis, K.S. 1981. External agents and knowledge utilization. In *Improving schools: Using what we know*, R. Lehming and M. Kane (eds.). Beverly Hills, CA: Sage Publications.

McKinney, Kay. 1993. *Improving math and science teaching*. Washington, D.C.: GPO.

McLaughlin, M.W. 1987. Learning from experience: Lessons from policy implementation. *Educational Evaluation and Policy Analysis* 9 (2).

Nelson, C.E., Roberts, J., Maederer, C., Wertheimer, B., and Johnson, B. 1987. The Utilization of social science information by policymakers. In *American Behavioral Scientist*, 30 (6), Sage Publications.

Smith, N.L., ed. 1981. *Federal efforts to develop new evaluation methods*. San Francisco: Jossey-Bass.

Smith, N.L., ed. 1982. *Field assessments of innovative evaluation methods*. San Francisco: Jossey-Bass.

Smith, N.L., ed. 1981. *New techniques for evaluation*. Beverly Hills, CA: Sage Publications.

Suarez, T.M., Montgomery, D.L., and Dania, S .T. 1989. Stakeholder II: The use of stakeholder analyses of anticipated outcomes in the design of summative evaluation. Paper presented at the meeting of the American Evaluation Association, San Francisco.

van de Vall, M. 1987. Data-based sociological practice. In *American Behavioral Scientist*, 30 (6), Sage Publications.

Weick, K. 1984. Small wins: Redefining the scale of social problems. *American Psychologist* 39, 40-61.

Weiss, C. 1972. Evaluation research. Englewood Cliffs, NJ: Prentice-Hall.

Wicker, A.W. 1985. Getting out of our conceptual ruts: Strategies for expanding conceptual frameworks. *American Psychologist*, 40 (10).

Yin, R.K. 1993. *Evaluation design: Breaking new ground*.

As I listen to these papers and discussions, it becomes quite clear that I represent another face of science education—that which calls itself informal science education and includes institutions such as natural history museums, science centers, children's museums, zoos, aquariums, botanical gardens, community centers, youth organizations (4-H, Girls Inc.), and even theme parks. Informal science education generally is not well connected to the educational research loop. However, (consciously or unconsciously) it both uses the result of educational research and seeks to avoid them. Let me make some connections and observations.

- Because informal science often is an open system—museum visitors (apart from highly directed school visitation) use museums as a recreational outlet as much or more than as an educational experience—museums have to look at the educational process differently from the way the formal system does. This has forced museums (or at lest the currently more successful ones) into serious front-end evaluation, or needs assessment. They are customer-driven, rather than driven by current research or even the availability of technology as delivery system, even though "research says" that museums must devise some other strategy for presenting that particular body of material. If a museum goes ahead simply on the basis of research studies and staff-initiated approaches, it runs the risk of giving a party to which no one comes. Therefore, museums are very selective about what they use of the enormous body of research that is being generated by NSF, intermediaries, and other providers of materials and ideas.

- Also, because most museum visitors are self-selected and not part of any curriculum or content module, most museums have learned that they must be very careful about defining desired outcomes. Measurements of content knowledge are unimportant to dangerous. The expectation that one several-hour visit to a science center or natural history museum will make or break a scientific career is preposterous. Rather, as in Hezel's discussion of goal-free or naturalistic evaluation, they are concerned with attitudes, and like Mark St. John, have become very good at ferreting out reactions to science as a way of thought and a legitimate area of interest and learning.

- Museums are experimenting with new techniques in data gathering. Webb mentions videotaping as a way of recording events and experiences. Museums do a great deal of that, sometimes even recognizing at the time filming is ongoing that the presence of a camera may cause behaviors and reactions that would be quite different were the camera not there. They do a great deal of eavesdropping, both surreptitiously and openly, following visitors to see what they do. In these respects, museums are able to be more creative than are researchers attempting to understand what goes on in a classroom.

- However, there are times when museums behave like the formal system and even are integrated into schools, and when museums serve as Dwyer's intermediaries for dissemination of ideas and materials into schools. For many years the Lawrence Hall of Science, a unit of California-Berkeley, has produced curriculum materials. These are marketed as GEMS—Great Explorations in Math and Science. These materials are tested extensively, and teacher workshops are convened to assist with their penetration of the classroom. A number of projects funded by the Howard Hughes Medical Institute are generating more materials tightly connected to reform curricu-

la—the University of Nebraska State Museum, the Oklahoma Museum of Natural History, and the Buffalo Museum of Science are deeply involved in this effort.

In conclusion, I suggest that there is a science education universe that calls itself informal science education. Sometimes it is responsive to the varied research being done through NSF, mainly when it sees that there is a clear utility to the NSF products. And just as often, this universe sees the research efforts as being unimportant or producing inapplicable results. Because museums are a growth industry, and because they are becoming increasingly sophisticated at knowing what happens in their exhibits and programs, it may be useful for evaluation of NSF research efforts to begin to include their impact on the informal sector.

I must say that I am very tempted by the previous presentations to tell you about what our Center does, but since I was always a good little student who did what the teacher said, instead I will do as asked and comment on the three papers I was asked to review. First of all, I too found them quite eclectic and interesting. I'll give you some general comments and then try to be specific.

One paper begins with the author characterizing the program he is considering. I found this very useful because it sets a context that I thought was needed for the suggestions and recommendations he makes, even though these may be quite generalizable. Two of the papers make specific evaluation suggestions, some of which were presented orally by the authors. The third paper was quite general. Two of the papers (not the same two) give very long lists of particular evaluation questions that might be asked. As the last speaker pointed out, there already is a tremendous problem in bounding the questions, let alone the data collection efforts. The two papers giving long lists did not deal successfully with this problem.

From my perspective—and I've been in and out of government—all the papers were written very much from an evaluator's point of view rather than from the point of view of the clients for the evaluation. The clients for NSF are researchers and educational systems and institutions; those who ultimately are going to ask evaluation questions are OMB and Congress. If NSF doesn't satisfy these oversight bodies, none of us will be sitting here 5 years from now.

A common theme advocated by these papers is the need for dissemination. That raises the question of what is worth disseminating and how this is decided. Webb's paper discusses an internal process of self-evaluation. He suggests using videos as a way of communicating, but does not address the issue of the researcher needing to decide what he or she wants to communicate. In fact, none of the papers addresses this issue in any detail.

Dr. Hezel made a point (quoted from someone else) that is well worth talking about: The kind of evaluation we are considering at this conference should be conceived as part of a system of self-renewal rather than as a yes/no decision-making paradigm. I commend all the papers because I think they are written in that spirit; it is a critically important point of view in considering alternative nonaccountability types of evaluation.

A second major point that he makes in his paper concerns the importance of dissemination. But I see very little in this paper that tells me an acceptable way of deciding how or why one would want to disseminate particular evaluation findings. Dissemination costs lots of money; let's not fool ourselves about this. I've always been amused at the funding curve that characterizes the Federal government and also private foundations that support research and development in education, as contrasted to that of private industry. For education, research and development receive the lion's share of funds, followed by program development, and trailed by dissemination or marketing. Because Federal agencies haven't learned that dissemination is very costly, the issue of what one chooses to disseminate has to be taken seriously.

Several minor points about the Hezel paper. He urges against nose counting, which I appreciate in the context of nontraditional evaluation. On the other hand, he also rightly points out that one must count whether minority populations, disadvantaged populations, and so on, are getting benefits. Therefore, I found myself in a little bit of a quandary as to whether the paper advocates nose counting or warns against it. This needs clarification. There also is a distinction, made early in the paper, between qualitative types of methods as being appropriate for non-traditional evaluations and quantita-

tive types of methods being appropriate for the usual sorts of outcome and impact assessments. In fact, later in the paper the author suggests using both quantitative and qualitative methods to address evaluation questions of both types. I agree with his later statement; the author should reconcile these two apparently conflicting positions.

Hezel also discusses the notion of tracing the intellectual origins of an innovation or program being evaluated. It would be extremely difficult for me to do so in my own work. I can't document how my synapses work. Yes, as researchers and evaluators, we add long lists of citations at the end of everything we write, whether it's a proposal or a paper, but where the intellectual ideas actually came from and how they were synthesized to give rise to a new project would not be easy to trace. Another question Hezel suggests asking concerns duration of the project and the difference between the proposed and actual duration of the project. I'm not sure what that would tell us, but perhaps Dr. Hezel could respond.

The next paper I want to comment on is Chris Dwyer's. (In my comments, I am moving from the most general to the most specific paper.) She discusses the idea of using intermediaries as key informants. This is a useful approach, provided the intermediaries are chosen appropriately. In my view, however, her list of criteria on selecting the intermediaries is missing the most important one, that is, whether the intermediary is knowledge-searching. Does it even operate in a context in which it needs R&D and evaluation knowledge? If so, what are its search mechanisms and the filters it uses for selecting what to act on? A good example is one that Dwyer actually gives, the National Diffusion Network (NDN), which uses a very particular kind of filtering device for deciding whether to disseminate information about a given program or not. If the desired evaluation (or filter) wasn't built into the program in the first place, it will never make it through NDN because it doesn't provide the data on which NDN bases its decision. Or to put it differently, NDN defines quality through impact data, while the program may have quite different criteria. Another question concerns how the intermediary deals with the information it acquires. If the intermediary is a knowledge-seeking kind of organization, if it has defined filters by which it judges the quality of research reports, research and development products, or whatever, and if it also has a process for acting on its searches and judgments, then I agree that including intermediaries, as key informants is one strategy among a number that could be considered.

I would not start in the way she suggests, however, I would do some retrospective analyses, namely look at the intermediaries and what knowledge they are actually using. That may raise similar problems to those I noted earlier with respect to Dr. Hezel's recommendation on tracing the origins of ideas. Perhaps one could start with some specific practice that looks as if it had been influenced by some assessed program, and then trace back where the practice came from. If the tracing involves an intermediary, the practice may have multiple origins. A good intermediary, one that is out there to improve practice, should be using multiple sources of information, not merely relying on a single project or program as the sole source for its information.

I found Norm Webb's paper very interesting and thoughtful. He placed his discussion in the context of a specific program, so that one could follow how he was relating his four major suggestions to NSF's Research on Teaching and Learning (RTL) Program. The evaluation matrix he suggests makes us aware of having to look for both the successes and the failures. Failure contributes to our knowledge as well as success; we tend to forget that. We tend to believe that only success is good, but that's not true in research or even in development. For example, we may develop a program that works in some setting, but when we find out it doesn't work in other settings, that's very

important information. Webb's matrix reminds us of this.

Let me comment on his specific suggestions. Regarding the retrograde analyses, I may have misunderstood what he intends, but I think they might focus too much on the internal process of a particular researcher or project. I would feel that's too narrow a net, unless combined with other strategies. If it is just one component of an evaluation, then I think it's an interesting suggestion. Something like that might be a piece of a larger-scale evaluation of an NSF program.

This particular suggestion reinforces the general impression I had of all the papers I reviewed, namely, that they appear to be written from an evaluation rather than from a policy perspective. For example, Webb conjectures about the reason for the many and varied kinds of projects in the RTL Program. Possibly, as he says, this has to do with all the client audiences, their needs, and all the different avenues to pursue. More likely, since this is a field-initiated program, and the peer review system being what it is, I suspect the eclectic nature of the RTL Program comes about as much through proposal pressure exerted by good people proposing the things they want to do as through a desire to meet client needs. The perspective of the evaluator of education R&D is different; we are concerned with the use of R&D products. So that's why I say this set of papers is written from the perspective of the evaluator rather than the real world of Federal agencies, but that's fine. I commend NSF for going outside its own concerns to get a different sort of perspective.

I've noted that I feel Webb's first suggestion is too narrow—just looking at NSF generating—its own further work through its principal investigators. The second suggestion, video documentation, made more sense to me in his oral presentation than when I read it. When I read it, it seemed more like PR than like evaluation. But orally, Webb made the point that, in the process of creating such a video, one would have to think about what it is that is important to disseminate. I think that's a very valid point, as I noted earlier. But I want to reemphasize that there has to be more widespread dissemination than just to one's peers; that is, to people who generate the research or the development products or who make judgments about what is worthy of dissemination.

The suggestion for cultural analysis of the research community is a wonderful idea, but I wonder whether it will be of interest to Congress. Consider the creation of a community of scholars that can engage in the kind of dialogue we are having this morning. This seems like a good thing. However, I am reminded of something that happened in the 70s when lots of money was being poured into graduate fellowships and traineeships in order to create a science infrastructure. All of a sudden, there were lots of young researchers asking for research money, and OMB said "Oh, we have created a monster. We cloned all these researchers and now we've got to feed them. This has to stop." And it did stop. Well, all right, I love the idea of studying the research community, but history makes me ask, "What is the Hill going to say?"

The fourth suggestion that Webb makes is on generalizability analysis. I have not had a chance to see the paper from Western Michigan, so I'm not precisely sure what it says about cluster analysis. This may be a better approach than a statistical one. Random sampling to deal with the great variety of projects funded by a program such as RTL does not strike me as appropriate. I would prefer groupings of projects that in some way reflect the approach taken, the problem addressed, etc. The groupings would have to be thought through very carefully. After grouping, one might select a representative subset of projects from each group for evaluation. If, for example, you had 200 projects and created 10 groups, you could select 3 out of each group for further study. I think that might be a better approach than random sampling.

Let me end my remarks by thanking NSF for the opportunity to participate in this stimulating conference and the audience for your attention. I look forward to the publication of all the papers.