NATIONAL SCIENCE FOUNDATION
4201 Wilson Boulevard
ARLINGTON, VIRGINIA 22230

OFFICE OF
INSPECTOR GENERAL

MEMORANDUM

DATE: May 14, 2004

TO: Dr. Judith Ramaley, Assistant Director
Directorate for Education and Human Resources

Deborah H. Cureton

FROM: Deborah H. Cureton
Associate Inspector General for Audit

SUBJECT: Audit of NSF's Math and Science Partnership Program
OIG Report Number 04-2-003

Attached please find the final report on our audit of NSF's Math and Science Partnership Program. We received your response and have revised the report as we deemed appropriate. We have also included your response, in full, as an appendix to this report.

In accordance with OMB Circular A-50, please furnish our office with a time-phased corrective action plan to address the report recommendations within 60 calendar days of the date of this report.

I would like to express my appreciation for the courtesies and assistance provided by your staff during the audit. If you have any questions, please contact Karen Scott or Jill Schamberger at (703) 292-7100.

Attachment

cc : [redacted]

Audit of NSF's
Math and Science Partnership Program

National Science Foundation
Office of Inspector General

May 14, 2004
OIG 04-2-003

Table of Contents

Introduction

Background

The National Science Foundation's (NSF) Math and Science Partnership program (MSP) began as a presidential initiative in 2001 and was authorized as a specific program by Congress in December 2002. The MSP program focuses on an overarching goal to improve elementary and secondary mathematics and science instruction for the purpose of strengthening preK-12 math and science education, improving student achievement, and reducing the achievement gaps among student populations. To accomplish this, the program awards funds to projects in which pre-kindergarten through 12th grade (preK-12) schools unite with institutions of higher education and other partners to undertake activities such as developing new curriculum and improving teachers' skills.

The development of partnerships among major stakeholders including school districts, higher education institutions, businesses, museums, and others, is an important feature of the projects funded by NSF's MSP program. The partnerships are expected, in the short-term, to promote the program's goals of providing all students access to challenging courses; increasing the quantity, quality, and diversity of preK-12 math and science teachers; and establishing results-oriented projects that implement evidence-based educational practices. Ultimately, the projects and the overall MSP program are expected to improve student achievement in math and science.

Under the MSP program, NSF funds three different types of MSP projects: comprehensive; targeted; and research, evaluation, and technical assistance (RETA). Comprehensive projects focus on change in both higher education institutions and in school districts to improve student achievement in math and science across the preK-12 continuum. Targeted projects focus on improving student achievement in a narrow range of grades or have a disciplinary focus in math and/or science. RETA projects enhance the capacity of the MSP comprehensive and targeted awards to achieve their goals through projects focusing on the development of research, evaluation, and dissemination tools.

Through early fiscal year (FY) 2004, NSF issued three solicitations for MSP comprehensive, targeted, and RETA project proposals. In two general rounds of funding, NSF made 61 MSP awards ranging

from 1 to 5 years and totaling $460 million. During the first round of MSP funding in late FY 2002, NSF awarded over $231 million for 7 comprehensive awards ($140 million), 16 targeted awards ($82 million), and 16 RETA awards ($9.5 million).[1] In FY 2003, NSF issued its second solicitation, and awarded over $228 million for 5 comprehensive ($143 million), 7 targeted ($61.3 million), and 10 RETA projects ($24 million). At the time of our audit in FY 2004, NSF had issued a third solicitation and was in the process of reviewing and awarding proposals.

Staff in NSF's Education and Human Resources Directorate, Office of the Assistant Director, manage the MSP program. The staff is comprised of a Senior Program Coordinator, six program officers, and a science education analyst.

Expectations for MSP Project and Program Evaluations

The MSP program represents a significant investment in improving student achievement in math and science. In order to determine whether the funded activities actually have an impact on student achievement, the legislation authorizing NSF's MSP program outlines expectations regarding evaluations of individual projects, as well as for the program as a whole. The legislation explicitly directs applications for this program to include "a description of how the partnership will assess its success."[2] In addition, the authorizing legislation directs NSF to evaluate the program, and, at a minimum, to use a common set of benchmarks and assessment tools to identify best practices and materials developed and demonstrated by the projects.

Concern about obtaining useful information from the project's evaluations is reflected in written comments by the House Committee on Science. The Committee Views from the House Science Committee on HR 1858, the precursor to the law authorizing the MSP program, states, "The Committee wants to ensure that the partnerships are closely evaluated…. The Committee stresses the importance of in-depth, quantitative assessments of the partnerships. The assessments should use common metrics to facilitate useful comparisons, and should

[1] Two of these first year awards, one comprehensive and one targeted, were co-funded by NSF and the U. S. Department of Education (ED). ED's contribution of $12.5 million towards these two awards is not included in the total dollars stated here. In addition, NSF originally made 17 targeted awards, but one award was phased out by mutual consent.

[2] Pub. L. No. 107-368 (2002).

measure quantitative factors and not just attitudinal changes. Assessment should be an integral part of partnership activities."[3]

While later MSP solicitations explicitly require the projects to have quantitative measurements and an independent evaluator, the first solicitation contained less direct evaluation requirements. Nevertheless, the first solicitation did state, "Each partnership must carefully plan project evaluation to guide the annual assessment of progress and to measure the impact of the effort." It further stated that the evaluation should provide evidence of the strengths and weaknesses of the effort being implemented, facilitating the partnership's understanding of what works. Finally, it stated that the evaluation should be designed to respond to the need to analyze both the qualitative and quantitative data to determine the effectiveness of the partnership in contributing to positive institutional changes and student academic outcomes.

Objectives, Scope, and Methodology

Because the ability to determine defined and measurable outcomes is so important to evaluating the ultimate success of the MSP program, the objectives of our audit were to:

- Determine whether the MSP comprehensive and targeted award recipients have effective evaluation processes that include measures of the impact of their intervention strategies on student achievement, and
- Determine whether NSF has a plan in place to evaluate the overall performance of the MSP program.

To accomplish these objectives, we focused on comprehensive and targeted awards made in the first year, FY 2002. At the time of our audit, these projects had been active for approximately one year, and the projects should have submitted to NSF a variety of reports including their strategic, evaluation, and implementation plans, and their initial annual progress reports. From this initial group of 23 MSP awards, we judgmentally selected a sample of 3 of the 7 comprehensive awards, and 6 of the 16 targeted awards, based on a cross-section of the award dollar amounts. We excluded RETA awards from our audit because their purpose is to provide assistance to the targeted and comprehensive projects, as well as to research specific issues in support of the projects.

[3] U.S. House of Representatives, Committee on Science, Committee Views on HR 4664, Mathematics and Science Education Partnerships, July 11, 2001.

We conducted a variety of work to determine whether the MSP comprehensive and targeted award recipients have effective evaluation processes that include measuring the impact of the intervention strategies and activities on student achievement. To develop an understanding of what constitutes an effective evaluation process, we reviewed evaluation literature and met with professionals in the education evaluation community. Recognizing that there are many approaches to evaluating education programs, we sought to develop a list of basic elements that would be key to an effective evaluation process. Relying on the advice of an education evaluation consultant,[4] our review of the initial MSP program solicitation, and evaluation literature including *The Program Evaluation Standards*[5] and an NSF-funded document entitled *The 2002 User Friendly Handbook for Project Evaluation*,[6] we identified nine basic elements that constitute an effective evaluation process.

To assess whether the nine projects selected in our sample had effective evaluation processes, we reviewed the various plans and reports submitted by the projects to NSF, including their proposals, strategic, implementation and evaluation plans, and their annual progress reports. We reviewed these plans and reports to determine whether the various projects contained the nine effective evaluation elements. Upon completing our analyses, we met with the respective program officers and discussed our assessment of each project in detail, to determine whether NSF was ensuring that, where necessary, the projects were taking steps to address the elements that were missing or needed improvements.

In order to determine whether NSF has a plan in place to evaluate the overall performance of the MSP program, we interviewed NSF program officials and reviewed relevant documents. We also interviewed Westat, NSF's contractor for the MSP monitoring system, and obtained information about planned data requirements and issues surrounding the implementation of this system. Finally, we met with experts in the field of education evaluation, including those from other federal agencies, to determine what are reasonable expectations for the content and timing of a program evaluation.

[4] RTI International.
[5] The American Evaluation Association, an international professional association of evaluators, developed these standards.
[6] Frechtling, J. (2002). *The 2002 User-Friendly Handbook for Project Evaluation.* Washington, DC: The National Science Foundation.

We conducted our audit work between July 2003 and February 2004, in accordance with generally accepted government auditing standards.

Results of Audit

Overall, we found that five of the nine FY 2002 Math and Science Partnership projects we reviewed had a process in place to effectively evaluate, define, and measure the impact of the intervention strategies, activities and outcomes on student achievement in math and science. While the remaining four projects did not address all the elements for an effective evaluation process, they could reasonably take the actions needed to implement such a process. In addition, NSF could better ensure that each MSP project has an effective evaluation process by providing award recipients with the elements of an evaluation framework that they could use to design their individual evaluations to measure the impact of their activities on student achievement.

Additionally, although NSF is taking steps to plan for an overall evaluation of the MSP program, it has not yet established milestones and deadlines for developing and implementing this evaluation. While experts do not specify a point in time by which a program evaluation should be designed and implemented, ideally an evaluation would be in place at the time a program is implemented. By delaying the program evaluation, NSF may be missing opportunities to obtain constructive feedback, as well as to implement midcourse improvements and corrections in the MSP projects and the overall program. Thus, NSF should place a high priority on designing and implementing this overall evaluation process, and develop specific milestones and timelines for accomplishing these tasks.

NSF Needs to Provide MSP Project Evaluation Guidance

Evaluations throughout a project's lifespan play an important role in understanding the activities and outcomes of that project and in assessing whether it is meeting expectations.[7] The evaluation process itself, as well as the data collected through that process, provide information that the project can use to make ongoing corrections, when necessary. Evaluations also provide the framework within which a project measures and reports on the ultimate outcomes and impacts of its activities and intervention strategies.

To assess whether the MSP projects had effective evaluation processes, we identified the following nine elements that provide a basic framework for conducting effective evaluations of the MSP program. Elements 1 through 3, and 7 through 9, are based upon *The Program Evaluation Standards* developed by the American Evaluation Association, which are recognized as standards for the evaluation profession. Elements 4 through 6 incorporate requirements contained in NSF's FY 2002 MSP solicitation.

1. **Award recipient has a conceptual model for the project.** Every project should start with a conceptual model to ensure that a common understanding about the project's structure, connections, and expected outcomes exists. Also, the conceptual model assists in focusing the evaluation design on the most critical program elements. The conceptual model should include project inputs, activities, short-term outcomes and long-term outcomes.

2. **Project has evaluation questions and developed and defined, measurable outcomes.** Evaluation questions build on the conceptual model and (1) identify key stakeholders and audiences, (2) formulate potential evaluation questions of interest to the stakeholders and audiences, (3) define outcomes in measurable terms, and (4) prioritize and refine questions.

3. **Project has identified an appropriate evaluation design that includes formative and summative evaluations of site-specific outcomes.** An appropriate evaluation design includes a methodological approach and data collection instruments and

[7] The same criteria apply to evaluating programs, defined as a collection of related projects. However, our purpose was to determine if individual MSP projects have effective evaluation processes in place. Performing an overall evaluation of the MSP program is addressed later in this report.

determines what will be studied and when. An evaluation design includes formative and summative evaluations.[8]

4. **Project has documented the commitment level of each partner.** Partnerships are an integral part of the MSP program, and the projects must document each partner's level of commitment in order to know how their involvement impacted the success of the project.

5. **Project will collect data from identified sources that measure the project activities and student achievement.** Documenting project activities and measuring student achievement are pieces of information necessary in determining the outcomes of projects funded by NSF's MSP program.

6. **Project will collect data from identified sources that measure the project's ability to increase and sustain its activities.** Increasing and sustaining activities are a central point in NSF's MSP program. Identifying the data sources ensures that appropriate data will be available to assess this.

7. **Project has objective implementation measures.** Before an evaluation of a project's outcomes or impact can be done, the project must be operating as planned. Objective measures of the implementation provide a means for the project to complete an early check to ensure that essential elements are in place and operating. Measuring the implementation also provides the project with information on the variations of how the interventions are carried out and the effects of these differences on the outcomes.

8. **Project has a plan for quantitative and qualitative analysis of the project data.** Quantitative and qualitative information that are appropriately and systematically analyzed ensure that the evaluation questions are effectively answered.

9. **Project will develop audience-specific reports.** Various stakeholders in a project have varying information needs, and audience-specific reports ensure that the project provides appropriate information for all stakeholders.

[8] Formative evaluations assess the ongoing project activities and provide information to monitor and improve the project. Summative evaluations assess the project's success in reaching its goals. For those projects with research activities at more than one site (for example, multiple schools), site-specific data should be collected.

Effective Evaluation Process

Five of nine FY 2002 MSP projects we reviewed addressed the nine elements for an effective evaluation process to measure the outcomes of their intervention strategies and activities, and to evaluate their impact on student achievement. Each of these projects identified evaluation questions that determine what measurement data will be collected. Also, each has an appropriate evaluation design that corresponds to and will evaluate outcomes that are specific to the project. For example, these projects specified the instruments they plan to use to evaluate their projects, as well as discussed designs, such as quasi-experimental designs, that will be used to determine the effects on student achievement.

Similarly, all five projects provided information on how the projects would measure the implementation of their project's approach and methodology during the course of the project. For example, all five will collect data to measure the involvement of major stakeholders in the partnership. One of these projects will measure the degree to which teachers successfully implement standards-based and inquiry-oriented science and math programs. Another project will measure the implementation of education practices through classroom observations conducted by project staff, teacher leaders, and an external evaluator.

Conversely, the remaining four projects in the sample lacked between three and nine of the nine elements for an effective evaluation process. For example, all four projects had not addressed the element for identifying appropriate evaluation designs. One of these four projects had not developed an evaluation plan for three of its five stated project goals, thereby limiting how it can evaluate the entire project. Another project had not provided a detailed evaluation plan to ensure that all of the project activities were in place and operating, again missing the opportunity to thoroughly evaluate the project. A third project did not organize its evaluation design in relation to specific interventions, benchmarks, and project outcomes, making it difficult to track and evaluate its activities.

These four projects also did not provide information as to how they would measure the implementation of their intervention strategies, and, therefore, risk not being able to assess the effectiveness of their strategies. For example, one project plans to recruit scientists to be "on call" to assist teachers but has no measures to determine how often teachers actually contacted the scientists or whether the scientists successfully provided assistance. Another project plans

to count the number of participants in a credential intern program for teachers. However, the project does not identify what steps it will take to implement outreach and recruitment activities for this program, and, therefore, will not be able to measure the effectiveness of these activities in increasing the quantity of teachers.

In addition, two of these four projects did not document their partners' commitment levels, as addressed in the MSP solicitation. Furthermore, these same two projects did not provide information as to how they will collect data related to student- and teacher-activities. For example, one project includes a goal to increase schools' capacity to provide challenging curriculum for every student, yet the project does not address how it will collect data to determine if this goal is met. Without collecting the information that is necessary to measure these types of activities, these projects will not be able to assess their own success.

Without all of the elements of an effective, ongoing evaluation process in place, NSF cannot ensure that the projects provide useful information about both the results of the activities and intervention strategies, and the effect or impact these activities have on student achievement. This information is also necessary for other stakeholders, including project participants, to learn from the outcomes of the projects. Foremost, the information is needed by educators in the math and science community, as they strive to understand what methods and strategies are most likely to improve student achievement and lessen achievement gaps.

Guidance on Evaluation Plans

NSF could better ensure the quality of the evaluation plans for all of its projects by requiring evaluation plans as part of the program solicitation and providing a framework of minimum evaluation factors that projects should include in their plans. However, NSF officials have stated that they prefer to leave the design up to the individual projects, and then rely on the professional judgment and past experience of its program officers to guide evaluation plans for the projects funded under the first MSP solicitation. This has resulted in inconsistent quality of the projects' evaluation designs and plans.

For example, in our discussions with the program officers regarding essential missing evaluation elements of particular projects, we received varied responses and opinions regarding approaches to project evaluation. While one program officer outlined the same

key elements for an effective evaluation process that we identified, another program officer stated that not all the elements we identified were necessary, although he agreed that all projects should have goals, objectives, benchmarks, and a series of measures to evaluate ultimate outcomes. A third program officer does not have any pre-established expectations for the evaluation plans, but rather relies on past experience with projects and evaluations to determine what should be included in an evaluation plan. This program officer stated that researchers do not have to provide detailed evaluation plans because the key elements are intrinsically built into the evaluation, and therefore, will be accomplished unless the project evaluators are totally inexperienced.

While each project has different activities and methodologies, and therefore different evaluation plans, it is nevertheless important that NSF ensure that the evaluation plans are effective in assessing the outcomes of each project. Accordingly, providing the projects with an evaluation framework comprised of the nine elements we identified would ensure that evaluation plans are held to a minimum set of standards, and ultimately would ensure that information is available to measure the outcomes and impacts of the activities funded. Also, this framework of minimum standards would help ensure that criteria are consistently applied across the various projects, and avoid the situation we found where four of the nine projects lacked an effective evaluation process. With quality evaluation plans, meaningful project results can be disseminated to improve educational practices as math and science educators strive for ways to improve student achievement.

Since the original MSP program solicitation in FY 2002 was developed, NSF has strengthened evaluation requirements for the projects in successive solicitations by explicitly requiring projects to have quantitative measurements and an independent evaluator. During the course of our audit, some program officers advised us that they have already begun to coordinate with project staff to include the missing evaluation elements in the sampled projects' evaluation plans. Such actions are an excellent start to ensuring that all projects have effective evaluation plans. However, as indicated by our review, NSF must ensure that all MSP projects provide measurable and useful information about their intervention strategies, activities, and outcomes.

Recommendation

We recommend that the Assistant Director, Education and Human Resources Directorate:

1-1) Implement procedures to ensure that the nine basic evaluation elements discussed in this report are required and documented in all current and future MSP project evaluation plans. For all current MSP projects, the NSF program officers should verify that the basic evaluation elements are included in the projects' evaluation plans, and where needed, work with the projects to address elements that are missing or need improvement, and for future MSP projects, ensure that these elements are included in any future solicitations.

Agency Response and OIG Comments

NSF agrees that appropriate overall guidance for project evaluation should be included in program solicitations, but does not believe that a framework of required evaluation elements is necessary. NSF notes that the MSP program is a large research and development (R and D) effort, and its projects do not necessarily involve mature work that lends itself to intervention-sensitive evaluation – what NSF refers to as the *context of justification*. NSF indicates that, as an R and D effort, the MSP program places a premium on innovation, and, as such, faces the challenges of evaluating projects that may lack maturity and possibly even the tools needed to measure important project components. Further, NSF believes our report overstates the shortcomings we identified in four of the MSP projects. The full text of NSF's response to our recommendation is included as an Appendix to this report.

We recognize and acknowledge that many of the MSP projects are innovative and may be difficult to measure and evaluate. We also recognize that NSF has funded additional projects under MSP-RETA to enable partnerships to better assess their work. Nevertheless, we believe all of these projects are capable of some level of measurement in all nine elements. As NSF and Westat, NSF's contractor, point out in the response to this report, while designing evaluations of R and D projects in the early stages is different from designing evaluations of more mature projects, "both are expected to meet the requirements for high quality evaluation, such as clearly stated questions; appropriate designs, sampling strategies and data analysis procedures; clear specification of data collection approaches, and use of reliable and valid instrumentation." These elements, applicable to both early R and D and mature projects, are included in the nine basic evaluation elements we identified, and four of the nine projects we reviewed lacked at least two of these above elements. For example, one project did not have clearly stated evaluation questions, nor did it have an appropriate evaluation design, both noted as necessary elements for a high quality evaluation of an R and D project in the

early stages. Additionally, the other five projects we reviewed were also R and D efforts in the early stages, and were able to address to some degree, all of the nine basic evaluation elements we identified. Further, the authorizing legislation for the MSP program included requirements for accountability through evaluation. NSF recognized this need for accountability and required in its solicitation that:

> …the evaluations should provide evidence of the strengths and weaknesses of the effort being implemented, facilitating the partnership's understanding of what works.

NSF also notes that one important component of the evaluation of the MSP program is the collection of data on student achievement. We agree, and found that the nine projects we reviewed had plans to collect data on student achievement. We have revised our report language to ensure that this point is clear. What is also important, however, is what causes that achievement to go up or down, and measuring the actual activities to accomplish improved student achievement is as important as measuring the student achievement itself.

The nine basic elements in our evaluation framework provide for the linkage between the strategies and activities, and their ultimate impact on student achievement. However, as discussed in the report, we identified shortcomings in the collection of other data related to student and teacher activities. For example, one project planned to measure the number of students enrolled in a new science curriculum, but had no plans to measure how many students successfully completed the curriculum units or how well the students did.

We reaffirm our recommendation.

NSF Needs to Make Evaluating the MSP Program a Priority

While evaluations play an important role at the project level, program-wide evaluations also provide critical opportunities to examine, at a global level, the usefulness of the activities and strategies employed by the individual projects that comprise a program. Experts do not specify a point in time by which evaluators must design and implement the program evaluation plan, and gather and analyze the data to effectively assess the outcomes of the program. However, some experts agree that, ideally, an evaluation plan would be implemented when a program is implemented, so as to capture baseline data, annual benchmark data, and identify and address issues affecting the program's outcomes.

The MSP legislation requires NSF to conduct this critical process. While NSF has taken some initial steps, it needs to document and formalize its plans. NSF has assigned responsibility for leading and managing this evaluation effort to its Division of Research, Evaluation and Communication (REC) within NSF's Education and Human Resources Directorate. Further, NSF is developing a request for proposal to contract with an outside consultant to design and conduct formative and summative evaluations of the MSP program.[9]

However, NSF has not yet formalized its plans for a program evaluation process, and has not established definitive timeframes and deadlines for implementing, conducting, and completing the process. More than two years have passed since NSF made the first MSP awards, and NSF is losing valuable opportunities to obtain constructive feedback on the individual 5-year projects and on the program as a whole, to identify and address programmatic issues, and to make needed mid-course corrections and improvements. Furthermore, delays in obtaining and analyzing data about the impact of the MSP program on student achievement may lead to delays in stating, with confidence, whether or not the program is achieving its goals. NSF is also missing opportunities to provide constructive input to successors of the MSP program.

[9] NSF is also in the process of establishing an on-line monitoring system that will collect data from all of the individual projects funded by the MSP program. NSF plans to collect this data on an on-going basis, and make this data available to the experts conducting the program evaluation.

Therefore, NSF must make planning for and formalizing the MSP evaluation process a higher priority.

Recommendation

We recommend that the Assistant Director, Education and Human Resources Directorate:

2-1) As soon as possible, develop and document a comprehensive management plan for evaluating the MSP program. The plan should specify milestones, target dates, and deliverables for accomplishing the evaluation process.

Agency Response and OIG Comments

NSF states that planning for the overall MSP program evaluation has progressed further than the audit report acknowledges. NSF expects to issue a solicitation for the program evaluation by May 2004, and award the contract by October 2004. NSF also notes it is developing the MSP Information Management System (MSP-IMS) which will aggregate data currently being collected from the MSP projects in annual progress reports, and will analyze common data elements.

We agree that NSF is making progress in developing the solicitation for the overall program evaluation, and in developing the data reporting modules for the MSP-IMS. Nevertheless, we note that the solicitation for the evaluation has not yet been issued, and the MSP-IMS is still in the planning and review stages, with implementation not expected until the end of this calendar year. As such, in delaying the development of a comprehensive plan for evaluating the MSP program, NSF is losing valuable opportunities to obtain timely data to assess and address programmatic issues.

This recommendation requires the timely development and documentation of NSF's plan for evaluating the MSP program. We reaffirm our recommendation.

In conclusion, both NSF and the OIG share the common goal of ensuring successful and high quality evaluation of NSF's Math and Science Partnership Program at both the project and the program levels. We recognize that research and development brings unique challenges to evaluation and measurement components, and we appreciate the Foundation's continuing efforts in this area.

Attachment A

Response to

Audit of NSF's
Math and Science Partnership Program

National Science Foundation
Directorate for Education and Human Resources (EHR)
Math and Science Partnership (MSP) Program

May 3, 2004

Prepared by
[redacted], [redacted], MSP/EHR,
with assistance from the MSP staff

Overview

We have reviewed this Audit and agree that formal, high quality, credible evaluation should be a high priority for NSF, the Directorate for Education and Human Resources (EHR), and the Math and Science Partnership (MSP) program, at both the project and program levels. The MSP program is a large research and development (R & D) effort with multiple responsibilities, and its evaluation must respond to both a *context of discovery* and a *context of justification*.[1]

Some aspects of MSP work implement strategies for which the field has a reasonable knowledge base emanating from existing scholarship or prior experience (e.g., with particular professional development strategies or particular classroom strategies). When the intervention or work to be done can be largely determined in advance of actual implementation, the work -- by its very nature – is more mature and, in most cases, lends itself to adequate assessment with existing measurement tools. This type of work is compatible with intervention-sensitive evaluation (*context of justification*). On the other hand, R & D efforts place a premium on *discovery*: being innovative in nature, they lack maturity and often lack even the tools needed to measure important components of their work.

Westat's Joy Frechtling, author of the NSF-funded publication *The 2002 User-Friendly Handbook for Project Evaluation*, states[2]

> R and D efforts, by their nature, are focused on discovering or applying knowledge to create something new or different. R and D explores new ideas or new approaches and applications of the knowledge base in the service of solving problems and creating new knowledge. Frequently, in such projects, all details of the "what and how" are not defined in the early stages and must emerge as the work develops. Designing evaluations of such efforts is quite different than designing evaluations of more mature projects. While both are expected to meet the requirements for high quality evaluation, such as clearly stated questions; appropriate designs, sampling strategies and data analysis procedures; clear specification of data collection approaches, and use of reliable and valid instrumentation, there are some important differences. And it is important to distinguish between evaluation plans that are technically flawed and evaluation plans that are incomplete because of the nature of the R and D process itself.

Because the Audit of NSF's Math and Science Partnership Program has its primary focus on evaluation planning in the *context of justification* – with little attention to the equally important but challenging *context of discovery* that is embedded through the MSP effort -- we do not agree with the first recommendation in the Audit Report to "[i]mplement procedures to ensure that the nine basic evaluation elements discussed in this report are required and documented in all current and future project evaluation plans." We do agree that appropriate *overall* guidance for project evaluation should be included in program solicitations. As the MSP program has itself gained experience and expanded its

knowledge base, each successive solicitation has progressively and appropriately strengthened the evaluation guidance communicated to the field in a way that acknowledges both the *context of justification* and the *context of discovery* that characterize the MSP program. The Audit Report acknowledges this progressive strengthening and we appreciate that recognition.

With regard to the overall MSP program evaluation and the second recommendation in the Audit Report, planning has progressed steadily and in more substantive ways than are acknowledged in the Audit Report. We expect the solicitation for overall program evaluation to be issued in May 2004, with a contract to be awarded by October 2004. In the interim, data essential to documenting and evaluating program and project progress ***are already being collected*** by MSP Partnership projects, with submission to NSF through annual progress reports. In addition, an externally contracted MSP Information System (MSP-IMS) provides for the aggregation of these data and the analyses of common data elements. The yet-to-be-awarded program evaluation contract will then further analyze these data and formulate findings and recommendations for the overall MSP program.

Components of MSP Evaluation Package

Evaluation in MSP follows a comprehensive and multi-faceted approach that is sensitive to the context of the program and its multiple responsibilities:

(1) Project-level evaluation that includes common data to be reported and that is otherwise specific to the goals of each funded Partnership;

(2) An externally contracted Information Management System (MSP-IMS) – already awarded under contract to Westat -- that provides support for the ***collection across all funded Partnerships of common quantitative and qualitative data, including student achievement data and teacher data***, analyses of these common data, and reporting that will inform and measure the progress of the program as a whole and of its individual projects.

(3) An overall program evaluation that will analyze the data collected, formulate findings and draw inferences from those analyses, add the finer granularity needed to inform understanding, and make program-level recommendations; and

(4) The *MSP-RETA* (research, evaluation and technical assistance) component of the program that consists of projects funded to develop new tools, measurement instruments, and the human capacity to inform evaluation of work that is currently evolving – and yet to evolve – as individual Partnerships move progressively into the R & D cycle. The funded *MSP-Net* project is an electronic community of practice that facilitates the sharing of tools, products, and findings to inform in a real-time mode the work of the Partnerships and to enhance their evidence base.

The Audit Report of the MSP program focuses almost exclusively on the first and third prongs of this overall package. The second prong of the package, the Information Management System (MSP-IMS), receives little attention. Yet, the MSP-IMS brings consistency to the overall MSP data effort through its ***common core of quantitative and qualitative data required of all Partnerships, including student and teacher data***. As such, it is an important component that mitigates some of the Audit Team's concerns, such as those about evaluation plans for two projects that "did not provide information as to how it will collect student- and teacher-related data, information that is necessary to measure outcomes such as increasing student achievement and increasing teacher quality, quantity, and diversity. Unless corrected, these projects will not have collected or analyzed information that is integral to assessing the overall success of the MSP program" (p. 10, Audit Report).

The initial MSP-IMS modules for data reporting have already been shared with Principal Investigators and are undergoing necessary review and approvals, including that required by the Office of Management and Budget, with actual implementation to commence before the end of the current calendar year. The important collection and evaluation of student achievement data will be discussed in greater detail below.

Context for MSP Program Evaluation

The MSP program is a large research and development (R & D) effort with multiple responsibilities. First, there is a responsibility for strong public accountability, especially in support of *No Child Left Behind* (NCLB), for (a) improved student achievement in mathematics and science, K-12, and (b) an improved K-12 teacher workforce in mathematics and science. Second, the MSP program has a responsibility to build the nation's capacity for improving mathematics and science education, including the knowledge base that contributes to effective educational practice at all levels of K-16 STEM education and the effective engagement of STEM disciplinary faculty in this work.

Some aspects of MSP work -- especially those that implement strategies for which the field has a reasonable knowledge base emanating from existing scholarship or prior experience -- are compatible with intervention-sensitive evaluation (*context of justification*). When the intervention or work to be done can be largely determined in advance of actual implementation, the work -- by its very nature – is more mature and, in most cases, lends itself to adequate assessment with existing measurement tools. R & D efforts, however, lack such maturity and often lack even the tools needed to measure important components of the work being done. This Audit Report heavily relies on a paradigm of evaluation that is intervention-sensitive at both the project and program levels, with relatively little acknowledgment of the important evaluation challenges that confront any R & D program.

Frechtling [communication of February 27, 2004] describes some of the unique challenges of evaluating an R & D effort:

- It may not be possible to completely define the complete evaluation plan early on when parts of the R and D effort are not themselves clearly defined. In some cases, a project may have four high level goals, but until all four goals are defined in more operational terms, strategies for attaining them, and the evidence for success identified, neither a comprehensive formative or summative plan can be created. In such a case, it may be possible to develop a complete evaluation plan for only one or two of the components and defer development of the rest until the program has been more completely specified. This is not really a problem with the evaluation plan, it is an issue that the project must address.
- With R and D efforts, it is sometimes the case that what needs to be measured cannot readily be accommodated with the tools that are at hand. Measuring teacher content knowledge is, somewhat surprisingly, an outcome that falls into this category. While there are a number of tools that purport to measure change in content knowledge, most instruments do not do so adequately. Another example is measuring the development of a partnership. While many studies have examined the notion of partnerships, we are still struggling with both the definition of the construct and the best ways to assess it.

  In such cases, R and D is needed in the evaluation as well as in the program. One solution would be to defer the R and D on the program side until the measurement tool for evaluation is more firmly in place. In most cases this is neither possible nor preferable, and the two types of R and D have to go hand in hand. Again, this is not a problem with a specific evaluation plan, but rather the state of the art; the tools and methodologies we have are not always adequate.

This view of the distinction between new and mature efforts is supported by other scholars, including Shadish[3] :

> Brand new programs have not yet had time to work out program conceptualization and implementation problems. . . . In addition, less background information and fewer past evaluations are likely to exist for new programs, so more work will have to done 'from scratch.' Well-established programs may be more ready for outcome evaluation, and they may have a greater wealth of information already available on them.

Haertel and Means[4] write that:

> With early-stage projects, it is important to know how the innovation plays out in real classrooms, and the evaluator needs to be alert to unintended interactions with features of the environment that program designers may not have taken into consideration. Providing useful feedback to program developers and developing understanding of project implementation in context—that is, how the elements of the innovation influence teacher and student behavior—will be paramount concerns at this stage. . . . Almost any approach produces good results in some settings with some kinds of supports. Before recommending particular approaches for broader implementation, we need a basis for understanding the range of contexts within which desired results are and are not likely to be forthcoming.

The MSP program, as well as the Partnership projects it has funded, has a strong emphasis on R & D. The evaluation of the MSP program is itself an R & D effort. In those domains where MSP program work can be characterized as more mature,

evaluation processes are also more mature. In domains that are more vested in R & D, the evaluation processes are evolving. In particular, there are critical challenges related to the existence of the very tools and instruments needed for assessing project process. The fourth prong of the MSP Evaluation Package, the MSP-RETA component, is therefore essential to an understanding of the logic of MSP and its R & D nature.

Even in their developmental stages, the tools and methodologies funded under MSP-RETA enable Partnerships to better assess the effects of their work and – absent other valid and reliable mechanisms – may be the best or perhaps even the *only* tools available. Projects funded by MSP-RETA include, for example, a *Longitudinal Design to Measure Effects of MSP Professional Development in Improving Quality of Instruction in Mathematics and Science Education* (Rolf Blank, Council of Chief State School Officers); *Design, Validation, and Dissemination of Measures of Content Knowledge for Teaching Mathematics* (Heather Hill, University of Michigan); *Assessing Teacher Learning about Science Learning* (Patrick Smith, Horizon Research, Inc.); *Alternative Approaches to Evaluating STEM Education Partnerships: A Review of Evaluation Methods and Application of an Interorganizational Model* (Gordon Kingsley, Georgia Tech Research Corporation); and *Adding Value to the Mathematics and Science Partnerships Evaluations* (Norman Webb, University of Wisconsin – Madison). The instruments, methodologies, and tools under development in MSP-RETA are piloted and field-tested at MSP projects and other sites. The expected contribution of MSP-RETA to the evaluation capacity of the MSP program is an important component in assessing the viability of the overall management plan for MSP evaluation and in an overall understanding of MSP program logic.

Data on Student Achievement

Because student achievement data are viewed as an especially important component of the evaluation of the MSP, they deserve special attention in this response to the Audit Report. Analysis of each of the 23 approved Strategic Plans for the FY 2002 cohort of Partnerships provides documentation that ***all*** Partnership projects have approved plans for collecting data on student achievement and a viable system for such data collection. Goal III-1b of the FY 2003 GPRA Annual Performance Goals for Strategic Outcomes related to People states:

NSF will significantly enhance the quality of K-12 mathematics and science education available to all students in Math and Science Partnership schools.

Performance indicators:

- *Evidence in the award portfolio of the infrastructure to support high quality programs addressing issues related to teacher workforce capacity, including preservice education and inservice professional development of math and science teachers as well as alternative routes into the profession (e.g., scientists and engineers becoming teachers).*

- *Evidence within Partnership schools systems of the infrastructure needed to improve math and science education and to measure improvement, i.e., the adoption of appropriate assessments of student achievement, as well as the initiation of the collection of achievement data that can be disaggregated by ethnicity, socioeconomic status, gender, etc.*

EHR, with assistance from Westat, compiled and assessed performance information related to this goal and reached a preliminary conclusion in summer 2003 that this goal was achieved for the FY 2002 cohort of Partnerships. In subsequent independent review in early September 2003, IBM (International Business Machines Corporation) Business Consulting Services then verified "the reliability of the processes NSF used to collect, process, maintain and report data for the goal and the analyses of the MSP proposals and strategic plans performed by NSF staff, external panels of reviewers, and Westsat." IBM also validated "that the Directorate of Education and Human Resources reached a reasonable conclusion that NSF achieved Goal III-B based on the quality of the performance information and analyses of the MSP program results to date."

Additional Comments on the Audit Report

We do not agree that the nine basic evaluation elements discussed in the Audit Report define an appropriate evaluation framework for an R & D effort, such as the MSP program. Within the framework of these nine basic elements, however, we would also suggest that the perceived shortcomings in the four MSP Partnership evaluation plans of focus -- when compared with these nine elements – have perhaps been overstated in the narrative of the Audit Report (p. 9). Supporting/additional information transmitted from the NSF Office of the Inspector General on April 7, 2004, about these projects shows, for example, that in one of the four projects, there is but a single element that is perceived by the Audit Team as "missing." When a project's evaluation plan is perceived as having part of a single element that is incomplete and ***its*** relative importance can be debated, tagging that project as having a less-than-effective evaluation process is arguably not very compelling. Comments/responses on the remaining three projects and other arguments aside, however, **perhaps the most important point to be made is that elements in a project's evaluation plan that are perceived by the Audit Team as "missing" are not always indicative of a flaw in *evaluation* planning, but rather of the *evolving project design* that emanates from its R & D nature.**

Nonetheless, the Audit Report and the accompanying Additional Information do identify some areas where project evaluation plans could be improved. The FY 2002 cohort of Partnerships that were the focus of the Audit Report are relatively young – about eighteen months old – and we will continue to work with them to strengthen their evaluation plans as their project designs mature.

Summary

We agree in principle with the Audit Report of NSF's Math and Science Partnership Program that MSP and other NSF program solicitations should provide ever-increasing

guidance about project evaluation, as our own knowledge base and that of the field expand and deepen. We are committed to making MSP program evaluation an R & D effort in itself. Each solicitation issued by the program has been successively stronger in articulating expectations for project evaluation. Program Officers continue to work with funded Partnerships, including the FY 2002 cohort that is the focus of this Audit Report, to strengthen their project designs and their evaluation plans. In addition, steady and substantive progress has been made on the overall external program evaluation, consonant with the R & D nature of the program. The MSP program is committed to a high quality evaluation effort. We appreciative the time and effort given by the Audit Team to improving our work.

Endnotes

[1] Terms from Charles S. Reichardt, Professor, University of Denver, *EHR Series on Scientific Research Methods in Education*, March 29, 2004.

[2] Joy Frechtling, Westat, private communication of February 27, 2004.

[3] Shadish, W. (1998-1999). Some Evaluation Questions. *Practical Assessment, Research and Evaluation*, 6, 3.

[4] Haertel, G. D., & Means, B. (in press). Evaluating Educational Technology. New York: Teachers College Press.