This document is being published online-only. It is being provided by NSF to supplement the information presented in the FY 2013 Budget Request to Congress and is not a formal component of the NSF Budget Request.

FY 2010-FY 2011 STEM WORKFORCE PRIORITY GOAL
FINAL REPORT
FEBRUARY 13, 2012

Overview

Context

The National Science Foundation (NSF) falls under the Government Performance and Results Act (GPRA) of 1993 and related performance reporting guidance issued by the Office of Management and Budget (OMB). Early in 2011 the GPRA laws were refined and additional guidance provided. The GPRA Modernization Act requires agencies to increase their alignment between strategic goals and annual plans, establish short-term (approximately two-year) Agency Priority Goals, mandates more frequent performance reporting, and outlines consequences for missed targets. NSF, as part of its strategic planning process (see NSF Strategic Plan), develops annual performance goals. NSF's Annual Performance Plans, which are published each February in the Budget Request to Congress, support the mission of the agency and guide it toward meeting its strategic objectives.

Goal Information

Per the GPRA Modernization Act, some annual goals are identified as Priority Goals. NSF had one Priority Goal in FY 2010-FY 2011, in the area of development of the science, technology, engineering, and mathematics (STEM) workforce. The 2011 Priority Goal states:

> *By the end of 2011, at least six major NSF science, technology, engineering and mathematics (STEM) workforce development programs at the graduate, postdoctoral, or early career level have evaluation and assessment systems providing findings enabling program re-design or consolidation for more strategic impact.*

In the FY 2011 Annual Performance Plan, the STEM Workforce Priority Goal was included as an annual GPRA goal under the NSF strategic goal "Transform the Frontiers" objective T-2, "Prepare and engage a diverse STEM workforce motivated to participate at the frontiers."

Methods

This goal involved 25 NSF workforce development funding programs at the graduate, professional or early career level. To ascertain progress against the goal, a theoretical evaluation continuum was developed at the outset of the process, under the assumption that the stages would occur sequentially along a progressive continuum. However, this assumption did not hold because the on-going programs involved in the performance goal were already established and at various stages of maturity. The stages were therefore assessed independently to determine progress. Definitions of key terms were agreed to and, for each program, evidence of progress within each stage was documented.

Results

A systematic assessment of results for the 25 programs was conducted using a 5-point rating scale (see Table 1). While the target of the Priority Goal was for "at least" six programs to meet the goal, 12 of the 25 programs reached the goal of having evaluation and assessment systems in place providing evidence and data to implement program improvements. Of the 13 programs that did not reach the target, four are relatively new and had not been active long enough to assemble sufficient data and evidence on which to base future program directions. For at least three of the 13 programs, planning or implementation of the first external evaluations is in progress.

Analysis

The Priority Goal was achieved. Substantial staff resources, time, and expertise were contributed by the Directorate for Education and Human Resources, the Performance Team in the NSF Budget Division, and the Office of Integrative Activities (OIA). Of particular note was the expertise in goal-setting and logic models provided by EHR and the new capacity for data mining to address program-specific questions developed through OIA.

Of the 25 Priority Goal programs, 12 support postdoctoral fellows or early-career researchers. Some of these programs are very small and have few resources committed to evaluation. A mini-retreat for NSF staff leaders of these programs was held on August 31, 2011. Program staff shared their individual logic models, or, in some cases, collaborated to build models. In addition, the group worked together to build a generic logic model for postdoctoral programs that could serve as an umbrella for individual programs. Because many of the postdoctoral programs are small, this effort could help in terms of resource sharing, assessment and determining best practice.

Since all 25 of the participating cohort of programs included in the Priority Goal demonstrated progress toward the target, with 12 reaching the target, NSF will monitor the participating programs informally throughout FY 2012 to encourage continued improvement and success. This will allow the committed programs to continue and deepen efforts both to bring a stronger orientation toward evidence as a basis for program improvement, and to enhance agency-wide collaboration in workforce development investments.

Participating Programs and Framework

Background

The Goal Leader for the STEM Workforce Priority Goal was the Assistant Director of the Education and Human Resources (EHR) Directorate, Dr. Joan Ferrini-Mundy. 25 NSF workforce development programs at the graduate, professional or early career level were identified and invited to participate in the Goal. All programs make awards via the merit review process. Of these programs, nearly half (11) are postdoctoral fellowship programs. Some of these programs are very small and have few resources committed to evaluation.

TABLE 1. Programs Involved in STEM Workforce Priority Goal

Directorate	Program
Cross-directorate	CAREER (Faculty Early Career Development Program)
BIO	PRFB: Postdoctoral Research Fellowships in Biology
ENG	GRDS: Graduate Research Diversity Supplements
GEO	EAR-PF: Earth Sciences Postdoctoral Fellowships Ocean Sciences Postdoctoral Fellowships OEDG: Opportunities for Enhancement of Diversity in Geosciences
MPS	ACC-F: American Competitiveness in Chemistry Fellowships AAPF: Astronomy & Astrophysics Postdoctoral Fellowships MSPRF: Mathematical Sciences Postdoctoral Research Fellowships
SBE	SBE Minority Postdoctoral Fellowships
OCI	CI TRaCS: Cyberinfrastructure Postdoctoral Fellowship
OISE	EAPSI: East Asia & Pacific Summer Institutes for U.S. Graduate Students IRES: International Research Experiences for Students PASI: Pan-American Advanced Studies Institutes IRFP: International Research Fellowship Program PIRE: Partnerships for International Research and Education
OPP	Polar Postdoctoral Fellowship Program
EHR	REESE: Research and Evaluation on Education in Science and Engineering AGEP: Alliances for Graduate Education and the Professoriate GRFP: Graduate Research Fellowship Program IGERT: Integrative Graduate Education and Research Traineeships Program GK-12: Graduate STEM Fellows in K-12 Education Program SMP: Science Masters Program NOYCE: Robert Noyce Teacher Scholarship Program SFS: Federal Cyber Service: Scholarship for Service Program

Framework for assessment and measurement

A theoretical model of program design and evaluation was developed to help provide a framework for assessment and measurement.

- Stage 1 (Baseline Data): Program has baseline information about workforce development (evidence found in baseline chart, solicitation, program description);
- Stage 2 (Goals and Theories of Action): Program has clear program goals and program theory of action of change (evidence found in baseline data and/or logic model);
- Stage 3 (Program Outcome Measures): Program has established metrics (evidence found in baseline data and/or logic model);
- Stage 4 (Performance Management Systems): Program has a performance management system in place (evidence found in a description of a performance management system appropriate to program);

- Stage 5 (Collected Data and Information): Program has sufficient data accumulated from the performance management system (evidence are data collected as part of performance management system); and
- Stage 6 (Evaluation and Assessment System): Program has evaluation and assessment system--uses evidence and data to implement program improvements (evidence is documentation of discussions and decisions made regarding the program using the data).

The assumption behind the theoretical model was that in an ideal state the stages would occur along a progressive continuum. However, since the on-going programs were entering the performance goal with activities already established and at various phases in their development and lifespan, for this report the stages were assessed independently to determine progress. As the year evolved, there was growing evidence that the continuum concept aligned with practices in the stronger programs, thus suggesting that the theoretical continuum was useful and validated as best practice, even though it was not fully in place, or falling into place out of sequence, for many of the existing programs.

Definition of key terms

- *Assessment and Evaluation*[1]: Assessment is defined as data-gathering strategies, analyses, and reporting processes that provide information that can be used to determine whether or not intended outcomes are being achieved.[2] Evaluation uses assessment information to support decisions on maintaining, changing, or discarding instructional or programmatic practices.[3]
- *Logic Model:* The Logic Model is a process tool that has been used for more than twenty years by program managers and evaluators to describe the intended effect of their programs. As described by the Foundation Coalition, a logic model describes logical linkages among program resources, activities, outputs, audiences, and short-, intermediate-, and long-term outcomes related to a specific problem or situation. Once a program has been described in terms of the logic model, critical measures of performance can be identified. A logic model forms the foundation for a programmatic assessment and evaluation system coupled with the six-stage theoretical process described above.
- *Performance Management Information System:* A working definition was developed by the Goal Leader, based on input from staff representing this set of diverse programs. A Performance Management Information System is a framework describing how an NSF program or unit uses data and information to understand and make decisions about their program. The systems used in program management at NSF are varied and individualized to each program's needs and capabilities. Likewise, data gathering systems for programs vary, and typically include elements such as locally held spreadsheets of data, post-panel materials for routine Division Director briefings, regular Principal Investigator (PI) surveys, systematic annual report analyses, NSF's Enterprise Information System (EIS) generated Committees of Visitors (COV) findings and recommendations, portfolio analyses, or external evaluation results.

  Ideally a Performance Management Information System should include a logic model and answer the following questions:
  1. What data or information is collected regularly?
  2. How are data (qualitative or quantitative) summarized or analyzed?

[1] Foundation Coalition definition, www.foundationcoalition.org.
[2] Gagne, R.M., L.J. Bridges, and W.W. Wagne. 1998. *Principles of Instructional Design.* Orlando, FL: Holt, Rinehart and Winston, Inc.
[3] Hanson, G., and B. Price. 1992. Academic Program Review. *In*: M. A. Wjitley, J. D. Porter, and R. H. Fenske (eds.). *The Primer for Institutional Research.* Tallahassee: Association for Institutional Research.

3. On what schedule are data reviewed (e.g., monthly, quarterly, annually, as necessary)? Is the review schedule linked to decision points (e.g., solicitation deadlines) or external assessments (e.g., COV meeting times or Advisory Committee meetings)?
4. How do the processes of data collection and review inform performance management of the program?

Results

At the close of the performance period, September 30, 2011, materials from each of the 25 programs were independently assessed by an expert against the stages of the evaluation continuum in the theoretical model described previously. The materials assessed were submitted by the programs or were available in NSF records. Each stage was examined separately to assess the program's progress towards the goal of developing "evaluation and assessment systems providing findings enabling program re-design or consolidation for more strategic impact."

Table 2 summarizes how the 25 participating programs were rated on a five-point scale in each stage of the process of building an evaluation and assessment system. On the rating scale, zero means no evidence of activity in a given stage; one=poor; two=started; three=good; four=substantial; five=excellent.

TABLE 2: Summary of Program Ratings on Six Stages

Program (names redacted)	Stage 1, Baseline Data	Stage 2, Goals and Theories of Action	Stage 3, Program Outcome Measures	Stage 4, Performance Mgmt. Systems	Stage 5, Collected Data and Information	Stage 6, Evaluation & Assessment System
1	**5**	**5**	**5**	**5**	**5**	**5**
2	**5**	**5**	**5**	**5**	**5**	**5**
3	**5**	**5**	**5**	**5**	**5**	**5**
4	**5**	**4**	**4**	**5**	**4**	**5**
5	**5**	**4**	**4**	**5**	**5**	**4**
6	**4**	**4**	**4**	**4**	**4**	**4**
7	**4**	**3**	**3**	**4**	**4**	**4**
8	**4**	**3**	**3**	**3**	**4**	**4**
9	**3**	**3**	**3**	**3**	**3**	**3**
10	**3**	**3**	**3**	**3**	**3**	**3**
11	**2**	**3**	**3**	**3**	**3**	**3**
12	**2**	**3**	**3**	**3**	**2**	**2**
13	3	3	3	3	3	0
14	2	3	3	3	2	0
15	4	0	0	4	4	0
16	3	2	2	3	2	0
17	3	3	3	0	0	0
18	3	3	3	0	0	0
19	3	3	3	0	0	0
20	3	0	0	3	3	0
21	2	3	3	0	0	0
22	2	2	2	0	0	0
23	3	0	0	3	0	0
24	2	2	2	0	0	0
25	2	0	0	0	2	0

Programs that reached or exceeded the target are listed in bold. Program names are omitted because the determinations were made using internal data.
For information on the documents and evidence used in this analysis, email performance@nsf.gov.

While the target was to ensure that six programs had an evaluation and assessment system in place, 12 programs out of the 25 reached the goal of having a system in place that used evidence and data to implement program improvements. These programs are listed in Table 3.

TABLE 3: Programs Reaching Target, their Focus and Managing Directorate

Program	Directorate	Focus
Alliances for Graduate Education and the Professoriate (AGEP)	EHR/SBE	Graduate
Faculty Early Career Development Program (CAREER)	EHR	Early Career
Fellowships for Transformative Computational Science using Cyberinfrastructure (CI-TraCS)	OCI	Postdoctoral
Earth Sciences Postdoctoral Fellowship (EAR-PF)	GEO	Postdoctoral
Graduate Research Fellowship Program (GRF)	EHR	Graduate
Integrative Graduate Education and Research Traineeship Program (IGERT)	EHR	Graduate
International Research Fellowship Program (IRFP)	OISE	Postdoctoral
Mathematical Sciences Postdoctoral Research Fellowships (MSPRF)	MPS	Postdoctoral
Robert Noyce Teacher Scholarship Program (NOYCE)	EHR	Postdoctoral
Opportunities for Enhancing Diversity in the Geosciences (OEDG)	GEO	Graduate
Postdoctoral Research Fellowship in Biology (PRFB)	BIO	Postdoctoral
Scholarship for Service/Cybercorps (SFS)	EHR	Graduate

The Six Highest Scoring Programs

The six highest scoring programs—AGEP, IGERT, NOYCE, OEDG, PRFB, and SFS--were distributed across three directorates. All of these programs are long-standing with more than a decade of experience in data collection, analysis and reporting. All have had at least one major external, summative evaluation, and they have employed a variety of evaluation methods including quasi-experimental designs, longitudinal studies, and mixed methods more than other, newer programs. Additionally, each of these programs seeks feedback loops and mines their data between summative evaluations, and some have used the Committee of Visitors mechanism as a way of making mid-stream adjustments.

Most of these top-ranking programs evolved with considerable Congressional direction, and there are dedicated staff members associated with these programs who are committed to implementing processes for evaluation and monitoring. In addition, funds for evaluation have been allocated for a longer period for the top programs than for many of the other participating programs. Given these patterns, it is not surprising that they should rank highest. Their experience has helped to inform the Priority Goal process and provided models to those staff who are less experienced and/or who work with smaller programs.

Major Accomplishments

Participating programs with mature evaluation and assessments systems provide findings that feed into the management and decision making process and lead to more strategic impact. Examples from four participating programs serve to illustrate the point.

- *Alliances for Graduate Education and the Professoriate (AGEP):* Project level evaluation results that were widely variable in quality and usefulness have led AGEP to recommend a new model for program-level evaluation based on a model used by several private funding organizations. The benefits are expected to be significant – providing technical assistance on formative evaluation for

each project and providing project level summative evaluation to standardize and improve the quality of the project level summative evaluations. Data trends from 1990 – 2007 for AGEP institutions and non-AGEP institutions show that AGEP institutions have enrolled and graduated more under-represented minority (URM) STEM graduate students compared to non-AGEP institutions by a factor of more than 2. The data indicate that the AGEP program may have been very good at identifying institutions that were already doing very well with URM recruitment and persistence, rather than stimulating increases beyond the existing trends in URM participation at these institutions. Discussions are underway about the implications for the future AGEP solicitations.

- *Robert Noyce Teacher Scholarship Program (NOYCE)*: The NOYCE program ensures that projects are collecting the necessary data and are familiar with monitoring and evaluation by holding new awardee sessions at each annual conference. They also launched a PI Guide in 2011 to draw on the expertise of experienced NOYCE PIs in guiding new projects. These efforts are in response to data collection indicating that projects are typically less successful in their first year in terms of recruitment. The current evaluation will study the relationships between the types of supports, activities and training that NOYCE recipients receive, the types of NOYCE recipients, and the recipients plans to go into and stay in teaching and leadership roles. In addition to informing the research community about effective practices, this will enable the program to emphasize specific types of support, activities or training in future solicitations.
- *Integrative Graduate Education and Research Traineeship Program (IGERT):* IGERT determined from its evaluation and assessment system that in the past most evaluations have been quantitative and were directed toward statistical inference on outcomes. Now IGERT plans to initiate a qualitative study to inform the STEM professional development community about best practices for crossing disciplinary boundaries and the role of institutional practices and policies. In FY 2013 they will develop a statement of work that highlights innovation. A pre-evaluation workshop is under consideration to build understanding of innovation in the context of graduate student education and training. In the future, the IGERT solicitation will include an educational research question that seeks to understand better the effective elements of institutional environment on the successful development of STEM Professionals at the doctoral level.
- *Opportunities for Enhancing Diversity in the Geosciences (OEDG):* The OEDG program uses an independent external contractor to perform periodic summative evaluations of the program to assess success in achieving goals. As part of its most recent summative evaluation, an expert panel was convened to review the annual project reports in order to document impacts, outcomes and make recommendations for program management. The panel recommended specific considerations and approaches that individual projects should undertake, and these recommendations were explicitly shared with new OEDG awardees at the annual PI meeting. The panel also recommended more rigorous data collection (particularly demographics for all participants as well as project impact on all participants, not just the target audience). These recommendations and possible new requirements are under discussion inside NSF and with awardees. The Panel also encouraged long-term tracking of project participants and the program is actively considering cost-effective mechanisms for doing so that comply with the Federal regulations regarding privacy.

Variation Among Programs

As Table 1 indicates, the quality of documentation varied even among the 12 programs that ranked on all six stages of the continuum. Some programs were well-documented and showed considerable evidence of using evaluation data to make decisions regarding the program. These programs were often at the end of a solicitation cycle, so considerable reflective work was in process, and this Priority Goal effort was able to capture that fully. Other programs were just getting started with a very ambitious longitudinal process and while they were able to describe intentions, data had not yet been gathered in many instances.

Timing of the Priority Goal in relationship to the lifecycle and maturity of the programs and the standard NSF rotation of program directors affected the ability of a program to be fully synchronized with the model and meet the target. It appears from the analysis that, for some programs, staff preferred to start at the first stages of the continuum and employ a linear process of building the system, while others tended to tackle aspects of all stages simultaneously but often less extensively and with lower levels of documentation.

Of the 13 programs that did not reach the target, four are relatively new and had not been active long enough to assemble sufficient data and evidence on which to base future program directions. For at least three of the 13 programs, planning or implementation of the first external evaluations is in progress.

Analysis, Next Steps, and Value Added Activities

The target for this Priority Goal was met and exceeded. The six programs scoring highest in the analysis have effective evaluation and assessment systems in place and are using the findings for program change and improvement. Evidence suggests they will continue to use these systems to make decisions related to these programs. The second tier of six programs also has assessment and evaluation systems in place and they have met their targets, but the programs and the performance management systems are less mature and need more time and experience to fully solidify. It is important to note that all 25 programs made progress towards achieving the goal.

Substantial staff resources, time, and expertise were committed through the Directorate for Education and Human Resources, the Performance Team in the NSF Budget Division, and the Office of Integrative Activities (OIA). Of particular note was the expertise in goal-setting and logic models provided by EHR and the new capacity for data mining to address program-specific questions developed through OIA.

Recommendation Regarding Next Steps

In another 12 months, NSF expects that the number of programs reaching all six stages of the model will grow and most programs appear headed to greater coherence. Because 12 programs reached the target and the rest of the participating cohort of programs demonstrated progress towards reaching the target, informal monitoring of this cohort should continue for another year to encourage progress, solidify preliminary success, and reap the results which are still latent or only partially realized. This will allow the committed cross-directorate team to continue and deepen its efforts both to bring a stronger orientation toward evidence as a basis for program improvement, and to enhance agency-wide collaboration in workforce development investments.

OIA Collaboration

In FY 2011, NSF's Office of Integrative Activities (OIA) began developing capabilities for NSF-wide data mining and analysis of available program information. The office collaborated with a number of Priority Goal programs to attempt to address specific questions formulated by the programs as part of the Priority Goal process. This new capacity for improved data mining of existing program documents provided substantial evidence for evaluative analysis of the programs participating in the Priority Goal activity. Given NSF's commitment to developing Foundation-wide evaluation capability, the OIA participation in this component of the Priority Goal was critical in helping to build needed resources.

Postdoctoral Programs Mini-Retreat –an Innovative Approach to Stakeholder Engagement

Early in the year, after completing a list of NSF programs that could be considered relevant to NSF's Priority Goal, the Goal Leader, Dr. Joan Ferrini-Mundy asked the leader of a long-standing grassroots committee of program officers who direct NSF's postdoctoral fellowship programs, Dr. Carter Kimsey, to convene the "Post Doc group" to discuss the suitability of including the postdoctoral fellowship programs in the Priority Goal. They agreed to join and participated actively in the Priority Goal. Some of the NSF postdoctoral fellowship and funding programs have been in existence since 1979-80 while other programs were started more recently. In some cases, program evaluations have been conducted. However, most of the postdoctoral fellowship programs are new and small, with fewer than ten fellowships awarded each year, and evaluation is done in a local, sometimes ad hoc, way.

Members of the postdoctoral fellowship subgroup had meetings to learn about logic models and undertook developing individual logic models for each program with the help of an expert consultant. The group convened a mini-retreat to share logic models and discuss the possibility and advisability of generating a generic logic model for postdoctoral fellowship programs at NSF. Such a model could be used to harmonize existing programs around common goals and address common problems across all

disciplines, assist in establishing new programs of postdoctoral fellowships, evaluate existing programs, and make improvements in the program management of existing programs. The postdoctoral group was successful in creating a generic logic model (see Figure 1). They also validated their logic model work in a workshop with Post Doc awardees at the National Postdoctoral Association (NPA) Appreciation Week in September 2011.

Figure 1.

Revised DRAFT LOGIC MODEL – Generic NSF Postdoctoral Fellowship

Program Goals: To diversify and strengthen the Nation's STEM research workforce and its leadership capacity, early career postdoctoral fellowships are given to individuals who choose research environments to foster their development. These fellowships spark creativity, nurture emergence of new interdisciplinary fields, enable discovery, and contribute to broader societal impacts.

Situation /Problem	Input(Resources Invested)	Activities /Training	Outcome Indicators	Short term Outcomes	Long term Outcomes
•Science is becoming more global and multi-disciplinary requiring greater vertical and horizontal integration; •Diversity is an issue in STEM disciplines, need to call for efforts to broaden participation • Need to develop independent scientific leaders to strengthen disciplines and keep the US competitive	•NSF funds and staff •Mentor/Sponsor-effective mentoring, time and resources •Institutional contributions – e.g. recognition of Fellows, lab space, computational resources, technical assistance, training, equipment, access to resources and benefits, etc. •Other – depending on partner and context •Required postdoctoral mentoring plans	•Independent research •Training activities •Reporting •Mentoring •Career Development •Communication •Broader Impacts	•Retention in the profession •Publications, patents, products, tools •Presentations, Poster sessions, knowledge sharing •Leadership as evidenced in things like organizing meetings; community outreach; grant proposal preparation; teaching; proposal review; peer review •Flexibility and adaptability demonstrated by Fellow, mentor, host and NSF	•Independent research agenda •Measureable impact on the science, the individual, the mentor and the institution •Critical mass in building professional community •Successful transition to next step in career •Success in subsequent research funding •NSF effective and efficient in support during fellowship	•Leadership in science research and professional advancement •Quality of science improves over time •Quality of public understanding of science grows •Achievements reflect positively on NSF - value added to field and society •NSF better able to partner and manage with flexibility and adaptability.

Assumptions: 1) Direct support to early career scientists to establish and pursue an independent research agenda yields greater creativity, innovation, discovery and leadership; 2)NSF awards to individuals are particularly prestigious and attract strong proposals from applicants with significant potential; 3)Sponsoring scientists will welcome Fellows who bring their own salary, research support, and state-of-the-art approaches; 4) Independent support is required to strengthen experimentation within disciplines, help young investigators navigate across disciplinary silos, and engage with broader societal impacts; and 5)Portability of the fellowship and flexibility in administration are key to fostering the success of early career scientists.