Framework for Evaluating Impacts of Broadening Participation Projects

Report from a National Science Foundation Workshop

The National Science Foundation

The Directorate for Education and Human Resources

The Division of Research on Learning in Formal and Informal Settings (DRL)

Framework for Evaluating Impacts of

Broadening Participation Projects

Report from a National Science Foundation Workshop

Co-Editors:
Beatriz Chu Clewell
Norman Fortenberry

Authors:
Fitzgerald Bramwell
Patricia B. Campbell
Beatriz Chu Clewell
Darnella Davis
Norman Fortenberry
Antonio García
Donna Nelson
Veronica G. Thomas
Adam Stoll

To cite: Clewell, B. & Fortenberry, N. (Eds.). (Jun 30, 2009). Framework for Evaluating Impacts of Broadening Participation Projects.

Prepared under Contract Number GS-10F-0482P, NSF Order Number DACS06D1421, Evaluation Activities Related to the Academic Competitiveness Council's Examination of STEM Education Programs.

30 June 2009

The National Science Foundation

The Directorate for Education and Human Resources

The Division of Research on Learning in Formal and Informal Settings (DRL)

TABLE OF CONTENTS

PART I: FOCUS ON NSF STEM BROADENING PARTICIPATION PROGRAMS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

We are extremely grateful to the authors who shared their considerable expertise in writing the sections of this *Framework for Evaluating Impacts of Broadening Participation*: Fitzgerald Bramwell (University of Kentucky), Patricia B. Campbell (Campbell-Kibler Associates), Darnella Davis (COSMOS Corporation), Antonio García (Arizona State University), Donna Nelson (University of Oklahoma), Veronica G. Thomas (Howard University), and Adam Stoll (Congressional Research Service). Very special thanks go to Norman Fortenberry (National Academy of Engineering) and Beatriz Chu Clewell (The Urban Institute) for their dedication to organizing and editing this publication, along with facilitating the NSF workshop on which it is based.

Thanks also to the external reviewers who provided many excellent suggestions for improvement and to those at NSF, especially Bernice Anderson and Frances Lawrenz, who were major contributors to this document. Additionally, many thanks for the editorial expertise provided by Pram Consulting Group Inc. and Guardians of Honor, LLC.

Elmima Johnson
Program Director

Division of Research on Learning in Formal and Informal Settings

INTRODUCTION

This report grew out of a workshop sponsored by the National Science Foundation (NSF) in Arlington, Virginia, on April 17-18, 2008. The two-day workshop, attended by approximately 60 participants, sought to develop and validate a strategy by which to assess the value of NSF's investment in broadening participation across all directorates and programs. Invited participants represented the following constituencies: NSF grantees, professional evaluators, and the policy community (which included representatives from Congress, the Office of Management and Budget [OMB], NSF staff, and staff from other federal agencies). Many of the workshop participants and other invited guests reconvened in December 2008 to hear about and discuss the progress being made in developing a report based on the April workshop. The December half-day meeting offered an opportunity to refine the ideas laid out by the authors engaged in developing the final document.

The workshop was structured around responding to two questions:

- What metrics should be used for project monitoring?
- What designs and indicators should be used for program evaluation?

The workshop resulted in providing information for NSF about what it should require for program monitoring and for program evaluation and advice and data gathering information relevant to awardees.

Speakers from NSF and OMB, respectively, discussed the NSF perspective on the importance of broadening participation and ongoing efforts of a recently established working group, and the OMB perspective on evaluating broadening participation programs (please see Appendix B for the workshop agenda). The major part of the workshop, however, was spent in small group parallel sessions to address the two questions above. The sessions were led by the authors of the various chapters contained in this volume.

The report incorporates the discussions in the plenary and small group sessions of the workshop within the parameters of the two framing questions above, but goes beyond the workshop in considering what might be NSF's approach to assessing its efforts to broaden participation across programs and directorates (Chapter One). "The Policy Context for NSF's Programs for Broadening Participation," by Norman Fortenberry, the second chapter in the report, lays out the policy context within which the NSF strategy for broadening participation has developed over the years, providing the background against which any discussion of assessment must take place.

The next chapter in this series, "Implications of the NSF Broader Impacts Statement for Broadening Participation: A Inclusive Strategy," by Nelson and Bramwell, comments on the way that the NSF broadening participation goal is expressed in NSF broader impacts statements and related activities. It gives recommendations for actions that will help to improve the way in which the Foundation goes about fulfilling its broadening participation goal with the data provided.

The fourth chapter, "Measuring Success and Effectiveness of NSF's Broadening Participation Programs: Suggested Monitoring Metrics and Evaluation Indicators," by Clewell, describes a recent effort by NSF to identify a broadening participation portfolio of funded programs and to classify these programs according to broadening participation goals. The paper then suggests appropriate monitoring metrics and indicators that could be used to evaluate the programs in this portfolio.

The remaining chapters discuss the evaluation of broadening participation efforts more broadly. They can be read and utilized independently. The fifth chapter, "Outcomes and Indicators Related to Broadening Participation," authored by Campbell, Thomas, and Stoll, provides an overview of outcomes and related indicators of success that might be used not only in evaluating science, technology, engineering, and mathematics (STEM) programs, but also in assessing Foundation-level efforts in internal and external areas. Its companion chapter by the same authors, "Evaluating Efforts to Broaden Participation in STEM Fields," focuses on issues of evaluation design, including appropriate evaluation designs for broadening participation-type projects.

As documented in the Fortenberry chapter, NSF's goal of broadening participation has been shaped by a variety of policy actions of the legislative and executive branches of government. Within the agency itself, policies articulated by the National Science Board (NSB) and the Committee on Equal Opportunity in Science and Engineering (CEOSE) have informed the NSF approach and strategy to address this goal, as referenced in major policy documents issued by NSF.[1] The recent NSF publication, *Broadening Participation at the National Science Foundation: A Framework for Action* (NSF, 2008), outlines the NSF-wide broadening participation plan. As such, it provides guidelines for broadening participation both externally and internally, through actions such as expanding the reviewer pool, training NSF staff and reviewers, enforcing accountability for NSF staff and principal investigators, communicating promising practices, and maintaining and monitoring a portfolio of relevant programs. Our report considers approaches to assessing the efficacy of these actions, with a primary focus on the evaluation of programs/projects that make up the broadening participation portfolio. We have chosen, nevertheless, in several of the chapters, to include a wider perspective on the task of evaluating these types of programs and activities, recognizing that the goal of broadening participation should be integral to all functions of the agency, transcending a discrete set of actions.

Beatriz Chu Clewell
Norman Fortenberry

Co-Editors

[1] See the Proposal and Award Policies and Procedures Guide (NSF 07-140), the *NSF Strategic Plan* (NSF 06-48) and the *NSF Budget Request*.

PART I: FOCUS ON NSF STEM BROADENING PARTICIPATION PROGRAMS

CHAPTER 1: SUMMARY OF THE WORKSHOP ON EVALUATION OF EFFORTS TO BROADEN PARTICIPATION IN STEM

Darnella Davis, Ed.D.
COSMOS Corporation

Antonio García, Ph.D.
Arizona State University
Hispanic Research Center

THE WORKSHOP CONTEXT

In April 2008, experts including NSF grantees, professional evaluators, and representatives from the policy community took part in a two-day workshop to develop strategies for demonstrating the value of the National Science Foundation's (NSF) investment in broadening participation across all programs and directorates. The gathering also included a reexamination of NSF's broader impact merit criteria for furthering broadening participation goals.

The two questions addressed during the April workshop were:

- What metrics should be used for project monitoring?
- What designs and indicators should be used for project evaluation?

Many of the workshop participants and other invited guests reconvened in December 2008 to hear about and discuss the progress being made in developing a report based on the April workshop. The December half-day meeting offered an opportunity to refine the ideas laid out by the authors engaged in developing the final document.

Presentations were organized by draft report section and covered by the respective author(s). Thus, discussion was roughly structured by topic, beginning with an overview of the draft report. Following the overview, there were presentations on the following topics:

- The Policy Context for NSF Programs for Broadening Participation,
- Measuring Success and Effectiveness in NSF's Broadening Participation Program: Monitoring Metrics and Evaluation Indicators; and

- Critical Issues Related to Indicators and Outcomes.

An additional presentation provided a context for these topics: Implications of the Broader Impacts Statement for Broadening Participation. The presentation was based in part on a paper prepared for the April workshop (Nelson and Bramwell, 2008).

Beverly Karplus Hartline, Associate Provost for Research and Dean of Graduate Studies, University of the District of Columbia (UDC), and member and former Chair of the Committee on Equal Opportunities in Science and Engineering (CEOSE), gave a response to the draft report.

Following the presentations, James H. Wyche, Division Director for the Division of Human Resources Development (HRD), facilitated an audience feedback session. Wanda Ward, Deputy Assistant Director for the Directorate of Education and Human Resources (EHR), then provided some final reflections before Bernice Anderson, EHR Senior Advisor, closed the proceedings.

SUMMARY OF WORKSHOP PRESENTATIONS AND AUDIENCE RESPONSES

Overview of Report

Norman Fortenberry, Center for the Advancement of Scholarship in Engineering, National Academy of Engineering, provided an overview of the draft report, explaining its goal of articulating useful metrics and broadening participation designs, and restating the two key questions addressed in the April 17-18, 2008 workshop. He also recalled that the aim of that workshop was to "develop and validate a strategy by which to assess NSF's investment in broadening participation across all directorates and programs."

The Policy Context for NSF's Programs for Broadening Participation

After the overview, Dr. Fortenberry transitioned into a presentation of the first topic which covers policy contexts for broadening participation, including those that fall within the legislative, executive, and judicial branches of government. He outlined the core values ideally embraced in programs and practices among for-profit organizations as well as NSF, the National Institutes of Health (NIH), nongovernmental organizations (NGOs), and the private sector. However, he noted that the impacts of broadening participation efforts by entities outside of NSF are not included in overall planning. In the case of NSF, he pointed to the role of CEOSE in providing guidance to NSF in its efforts to serve the public. He also noted the guidance on inclusiveness articulated in NSF's long-term plans:

Core value from the NSF strategic plan FY 2006-2011.

> *Broadly Inclusive:* Seeking and accommodating contributions from all sources while reaching out especially to groups that have been underrepresented; serving scientists, engineers, educators, students and the public across the nation; and exploring every opportunity for partnerships, both nationally and internationally.

Dr. Fortenberry mentioned implied metrics for gauging broadening participation impacts including participation (absolute or relative), impacts, and products. He noted that most public institutions address some of these metrics.

Dr. Fortenberry discussed a number of key policy statements, acts, and other directives containing language that sets the contexts that the different branches of government use in shaping broadening participation policy. The legislative branch focuses on authorizations and appropriations to higher education with foci beyond the top 50 universities, states that do not receive high amounts of federal research dollars, and community colleges. The executive branch has given orders that mandate efforts to strengthen Historically Black Colleges and Universities (HBCUs), Hispanic Serving Institutions (HSIs), and Tribal Colleges. The judicial branch has steered NSF programs from a focus on individuals based on their demographics to national outcomes to be achieved. Dr. Fortenberry concluded by enumerating some of the metrics that are emerging due to these policy perspectives as a way to gauge broadening participation. The metrics include rates of participation, indicators of impacts of institutional policies and practices, and measures of productivity in academic and professional products.

In sum: Emerging metrics are clarifying paths to achieving broadening participation in complex policy environments.

Measuring Success and Effectiveness in NSF's Broadening Participation Program: Monitoring Metrics and Evaluation Indicators

Patricia Campbell, Campbell-Kibler Associates, provided an overview of topic two, which was prepared by Beatriz Clewell of the Urban Institute. The second topic focused on NSF programs and their broadening participation strategies and it also highlighted appropriate metrics and indicators for monitoring and evaluating their effectiveness.

Two main types of broadening participation guided the discussion: individual and institutional. Given these parameters, two manners of capturing broadening participation data were outlined. The first covers monitoring metrics which capture short-term data such as stated goals, baseline data, or follow-up data. A second is that evaluation, which normally develops research questions and impact indicators, is longer term, and is situated within broader program-level goals. Ideally, the results of these data collection efforts are used by policymakers, funders, individual projects, researchers, and the practitioner community.

Recommendations for measuring success include:

- Collect common or uniform broadening participation data,
- Ask if programs serve proportional or representative groups; and
- Check if positive outcomes are equally distributed.

The interactive discussion thread included the following points:

- There is a structural challenge to having all data in the same format in that data might be; (a) required, (b) useful to have, or (c) perceived as beyond the interest of staff assigned to collect data that strictly adhere to a uniform standard.
- The agency and the field should work together to reduce the data collection burdens of each site in terms of required monitoring or evaluation.
- Additionally, evaluators may make recommendations to monitors in terms of capturing baseline data. Establishing baseline data is so important that evaluators should be engaged from the inception of a project.

In sum: Both monitoring and evaluation strategies can be refined to better gauge the progress and success of NSF's broadening participation program.

Critical Issues Related to Indicators and Outcomes and Evaluation Designs/Strategies

Veronica G. Thompson, Howard University Professor of Human Development, gave an overview of topic three, which argues that success is measured at multiple levels and important distinctions must be made among inputs, outputs, process, and outcomes. Inputs are colloquially defined as "What do we invest?" Outputs are "What do we do or who is served?" The process should be about tracking the implementation's alignment with the original intentions. Also, outcomes are not to be confused with process.

Dr. Thompson then defined the measures of success at multiple levels as:

- Access to the benefits of science, technology, engineering, and mathematics (STEM) knowledge,
- Access to STEM knowledge,
- Studying STEM,
- Working in STEM areas; and
- Generating knowledge.

Other considerations for fairly presenting data include:

- Parity as a range (e.g., 10-15 percent),
- Parity as more participation overall,
- Discipline or field size to which the definition applies; and
- Integrating qualitative indicators and transforming perspectives, (e.g., broader, more inclusive, diverse perspectives, or looking beyond numbers to policies).

For professional development, success may be seen at three levels: individual, institutional, or Foundation.

- At the individual level, the measures include participation, retention, persistence, success, experience, and attitudes.
- At the institutional level, other measures cover staff, policies, accountability, monitoring, and collaboration.
- At the Foundation level, still more measures address information about broadening participation, review policies, diversity of personnel, funding levels, knowledge gains, and strategies.

A number of questions and comments signaled some of the problems to be overcome in achieving a fair evaluation. One question is, "What do you use if the evaluator is not there at the start of the program? For example, what are the influences of prior experiences for students coming into an international program?"

Another possible solution would be the use of a retrospective design to address the challenges posed by an evaluation that does not begin at a project's inception. Or, adding questions as you learn more would be acceptable, as would the use of critical incidence. To address the challenge of isolating the current initiative, which is the main problem, it would be prudent to set up a good comparison group. In this respect, the Alliances for Graduate Education and the Professoriate (AGEP)[1] and the Louis Stokes Alliances for Minority Participation (LSAMP)[2] programs are teasing out multiple types of designs to determine impact.

The caution is that one needs to be careful about the difference between evaluation and research. One assertion is that the most one can get from evaluation is a preponderance of evidence and, therefore, triangulation is important.

Another point focused on the ability to make mid-course corrections. A point of clarification is that there is a distinction between mid-course corrections based on data versus a trial and error approach, and any corrections should be documented.

A more detailed discussion on how to make corrections that change the level of intervention was made in reference to medical care. One questioner asked, "Does it make sense to tweak dosage?" The reply to this was that intensity matters. An example of this is the American Association for the Advancement of Science's (AAAS) work for Innovative Technology Experiences for Students and Teachers (ITEST). Further discussion on this issue included a caution that there is a risk of confounding self-selection and dosage.

---

[1] Conducted by The Urban Institute, this evaluation report can be found at www.urban.org. Please see the Reference section of this chapter for the report links.

[2] Conducted by The Urban Institute, the evaluation reports can be found at www.urban.org. Please see the Reference section of this chapter for the report links.

Another question surfaced regarding whether the collaboration was being evaluated. If so, for informal science education (ISE), Randi Korn's chapter in *Framework for Evaluating Impacts on ISE Projects* (2008) was offered as a good reference.

Social network analysis was mentioned as having intriguing applications that can be joined with hierarchical linear modeling (HLM). Also, Mary Bucholtz currently has NSF funding for a study (The Role of Social Interaction in the Development of Scientist Identities and the Retention of Undergraduate Women in Science Majors) that should be of interest. In addition, NSF funded a retrospective study of collaboration covering computer science. And an analysis found that when more collaboration existed fewer publications were produced. However, it was noted that this finding does not directly establish a causal link between collaborations and publication rates.

Subsequently, there was a question about whether anyone is studying the interaction between research and evaluation. The interest was in whether within NSF's Math and Science Partnership Program Evaluation (MSP-PE), the stronger partner compromises the other. This prompted another set of questions: "First, what's the outcome of collaboration? What's the claim in terms of value added? Second, what's the process in terms of social network analysis as to how well collaboration works?" The point here is that studying collaborations and their outcomes constitutes a legitimate area of research.

Another participant noted that in an international social network initiative, what happened over time—as people spread—is not being evaluated. It was offered that in the future, the research community needs to measure large data sets that no one is currently tackling.

Pulling back to the underlying question about taking corrective actions, another participant asked "Which elements are working or not? How much control is there over variables? What actions lead to outcomes? How does one distinguish between research and evaluation?" One response is the idea of developing a center that merges research and evaluation while engaging graduate students, undergrads, and faculty. But another participant commented that the American Evaluation Association (AEA) is suggesting that the evaluation community stop trying to make a distinction between the two pursuits and just get more rigor in evaluation and more context in research. Another participant cautioned that the point is not to lose sight of the need to talk about the quality of collaborations, that it is important to measure the extent to which they are good or bad.

To clarify terms, Dr. Campbell offered the analogy that research is the dog, treatment is the tail. While in evaluation, the intervention is the dog and evaluation is the tail. Thus, the difference is in control. Evaluation doesn't wave the dog.

A useful reference on longitudinal tracking (Bailey, 2008) was offered during the discussion. In response, a participant observed that NSF now tracks minorities and women more than white males, and perhaps NSF projects should track all subgroups. However, it was noted that there are constraints due to confidentiality. Still, there was acknowledgement that evaluators can add questions for all groups to respond to voluntarily.

A new idea was posed relating to obtaining qualitative versus quantitative data. The question was "Why not use life histories?" In response, one suggestion was that a project should start with quantitative data and then go from there, using comparison groups and keeping in mind the hierarchy of research methods represented by the pyramid with experimental designs at the apex, followed by quasi-experimental, and then other designs.

When selecting designs, some checks include:

- Appropriateness of fit,
- Timing,
- Balance between level of evaluation and level of intervention,
- Level of evidence; and
- Strength of rivals.

Case studies can be used as a summative evaluation tool (Yin, 2009). One way is for a case study to document the outcomes of interest, which may be quantitative or qualitative or both. Another way is to address the attribution issue by explaining cause and effect relationships or enriching their understanding. Such insights can go beyond what can be discerned by using experiments or quasi-experiments alone, although the case studies cannot establish the cause-effect relationships with the same certainty as these other methods. In this sense, case studies also serve as a strong partner in complementing other methods as part of a mixed methods study.

Evaluation and Broader Impacts

After these presentations, an NSF program director referenced the contributions of Donna Nelson and Fitzgerald Bramwell's work in facilitating the April workshop, noting EHR's objective of collecting broadening participation data under all of the broader impact areas. In that context, NSF's merit criteria are the only places where awardees are required to report on broadening participation, although people can collect broadening participation data for the other four categories in terms of diversity, equality, and accessibility. However, with additional questions come additional costs and NSF must find funding to document and assess:

- Community outreach and dissemination,
- Integrating research and training,
- Building infrastructure,
- Potential societal benefits of human resource development,
- Reaching diverse media; and
- Encouraging use of research by diverse groups.

In commenting on the workshop, another NSF staff member, Dr. Fae Korsmo, discussed NSF's efforts to redefine its objectives through self-examination and its development of seven action items. She stressed the need to make the results of NSF's self-assessment accessible beyond the education community, paying special attention to the use of definitions and jargon. She mentioned that two studies are looking at broadening participation impacts. These efforts benefit from reviews of Committee of Visitors (COV)[3] reports and reviews conducted by AAAS staff. She also stated that NSF is open to redefining the broadening participation portfolio.

A participant then asked, "What are specific objects or special plans in regard to NSF's overall goals?" The response given was that NSF conducted a Foundation-wide survey that yielded 1,200 accomplishments. For 50 percent of these items, respondents checked that they related to "broadening participation" and explained what they had done.

The remaining question during this portion of the meeting summed up the challenges in evaluation: "Still, what's the best use of evaluation efforts, and at what level?"

[3] The NSF relies on the judgment of external experts to maintain high standards of program management, to provide advice for continuous improvement of NSF performance, and to ensure openness to the research and education community served by the Foundation.

Figure 1.1. Leveraging Broader Impacts

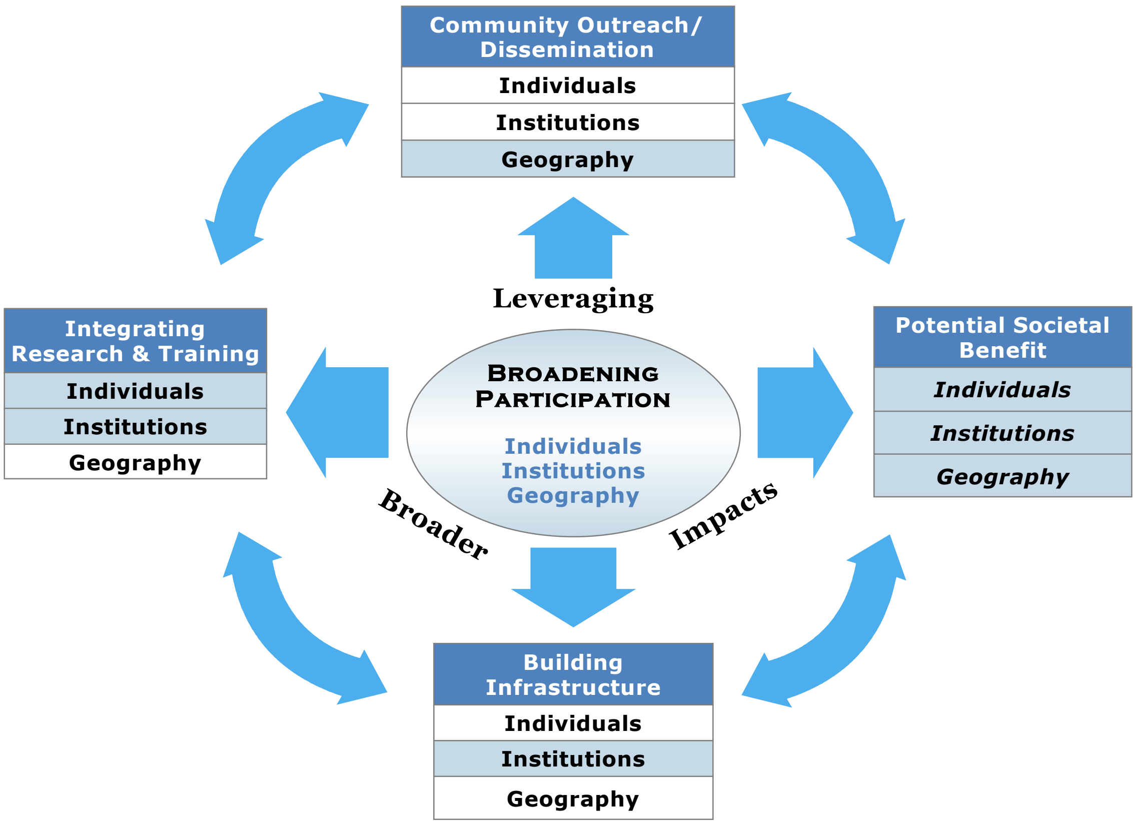


Figure 1.1. Contains a graphic depiction, Leveraging Broader Impacts, which illustrates the flow of potential broadening participation influences (from *Addressing Broadening Participation within the NSF Broader Impacts Category*, a presentation by Johnson and Anderson; based in part on Nelson and Bramwell, April 2008).

Beverly Karplus Hartline, Associate Provost for Research and Dean of Graduate Studies, University of the District of Columbia (UDC), and former Chair, CEOSE, was surprised to learn that the 2005 CEOSE report covering a decade of NSF activities showed that not much had changed over the years. For example, since the 1980s, except for increases in the number of white women, no other underrepresented group has moved toward parity. The report showed data with a positive slope, but noted that the populations of interest have also grown in magnitude.

Therefore, CEOSE is pushing for accountability, i.e., broadening participation *prima inter parus*. In fact, CEOSE is recommending that broadening participation be rewarded. Dr. Hartline personally applauds the workshops' efforts and would like to get the report out to those who might not otherwise be interested. Specifically, she would like to create more awareness that diversity is not the antithesis of excellence, nor is it an either/or choice. She asked "How can evaluation in general and this effort in particular help catalyze epiphanies to expand the ranks of the converted?"

The following are CEOSE's recommendations for broadening participation in research projects:

- There should be a FastLane reporting template that lists which underrepresented individuals or institutions have been impacted,
- Data should be disaggregated to capture important insights,
- NSF should acknowledge that impacts may be different for different demographic groups; and
- NSF should note that there are differential pathways for arriving at STEM, including various actions for encouraging student participation in STEM disciplines.

The point is to give attention to dimensions that are too often overlooked. This entails capturing the differential pathways to advancement at the faculty level, and at departmental, college, institutional, project, and program leadership and advisory levels.

The caution is over the importance of being critical of evaluations and their quality, paying attention to the credentials, quality, and performance of the various purveyors of evaluation. The question is "How can what NSF expects and requires be aligned with what its performing institutions are doing in assessment and accountability?" In this regard, another recommendation is to communicate issues about small numbers and privacy when, for example, there are fewer than 25 subjects and confidentiality is more easily compromised. It was noted that there is awareness of these issues. For example, for the redesign of their survey of earned doctorates, the Division of Science Resources Statistics (SRS) website is taking comments.

In sum: CEOSE supports NSF's current efforts to give greater attention to accountability and broadening participation while stressing the urgency of mainstreaming broadening participation transformations.

AUDIENCE FEEDBACK

Audience feedback was initiated with a provocative question: "What's between baseline and outcomes? What do we think is missing?" The audience was encouraged to identify other questions to be addressed in future evaluation dialogues.

A series of questions were then posed by another participant: "Are we happy with our outcomes? Where are the impacts? Why is there so much stasis when we should be seeing increases?"

In response to another participant's concern that NSF disciplinary heads were not present at the workshop, Dr. Ward noted that EHR is present as a leader and pushing all directorates and policies on broadening participation. It was pointed out that there is a small working group from NIH, NSF, and Research I universities asking how to engage. Additionally, there also is the Capacity Building in Evaluation effort.

Another set of research-focused questions resulting from the evaluation discussions were posed:

- What are cutting edge programs?
- What's going on elsewhere, internationally?
- What's the best way to scale up things that do work?
- What are the strengths—where does culture fit in?
- How does one jump start STEM in underrepresented communities?
- How does one use the advances in health technologies elsewhere?

A final comment: "There's a need for prioritization of the research and evaluation agenda, and that should drive the effort."

In closing, Dr. Ward commented that:

Evaluation has been a hallmark activity in EHR since 1992. The framework document is a major contribution to current and future evaluation practice, helping NSF and the field to improve both project and program evaluations. The recommendations are clear that an evaluation framework for broadening participation must employ multiple methods to provide guidance for continuous improvement in implementation and to determine the quality and impact of investments in promoting diversity, equity, and accessibility in STEM education and workforce development. Additionally, more longitudinal studies of individual and institutional performance should be conducted to assess investment returns.

REFERENCES

Bailey, T. (2008, February). Beyond traditional college: The role of community colleges, career and technical postsecondary education in preparing a globally competitive work force. *The New Role of Higher Education Attainment in Global Competitiveness and Income Opportunity: Implications for National Policy*, 23(1), 25–30.

Bronstein, M. (2007, December). *Analysis of Committee of Visitors (COV) Reports: 2004-2007*. Arlington, VA: National Science Foundation.

Clewell, B. C. & Campbell, P. B. (2008, January). *Building Evaluation Capacity: Guide I Designing A Cross Project Evaluation*. Washington, DC: The Urban Institute. Retrieved from http://www.urban.org/UploadedPDF/411651_guide1.pdf

Clewell, B.C., Cosentino de Cohen, C., Tsui, L., Forcier, L., Gao, E., Young, N., Deterding, N. & West, C. (2005, November). *Final Report: Evaluation of the National Science Foundation: Louis Stokes Alliances for Minority Participation Program*. Washington, DC: The Urban Institute. Retrieved from http://www.urban.org/UploadedPDF/411301_LSAMP_report_appen.pdf

Clewell, B. C., Cosentino de Cohen, C., Tsui, L. & Deterding, N. (2006). *Revitalizing the Nation's Talent Pool in STEM*. Washington, DC: The Urban Institute. Retrieved from http://www.urban.org/UploadedPDF/311299_revitalizing_stem.pdf

Cosentino de Cohen, C. (2006, March). *Building the Nation's Scientific Capacity: Evidence from the Louis Stokes Alliances for Minority Participation Program (LSAMP).* Washington, DC: The Urban Institute. Retrieved from http://www.urban.org/UploadedPDF/412231-building-the-nation.pdf

Committee on Equal Opportunities in Science and Engineering. (2004, December). *Broadening Participation in American's Science and Engineering Workforce, The 1994-2003 Decennial and 2004 Biennial Report to Congress*. Arlington, VA: CEOSE (CEOSE 04-01).

Johnson, E. & Anderson, B. (2008, April). *Addressing Broadening Participation within the NSF Broader Impacts Category.* Presentation for the National Science Foundation Workshop: Evaluation of Efforts to Broaden Participation in STEM, Arlington, VA.

Korn, R. (2008, March). Evaluating collaborations. In Friedman, A. (Ed.). *Framework for Evaluating Impacts of Informal Science Education* (pp. 87-98). Retrieved from http://informalscience.org/evaluations/eval_framework.pdf

National Science Foundation (2006). *Investing in America's Future: Strategic Plan FY 2006-2011*. Arlington, VA: National Science Foundation (NSF 06-48).

Nelson, D. and Fitzgerald B. (2008, April). *Evaluation of Efforts to Broaden Participation in STEM*. Presentation for the National Science Foundation Workshop: Evaluation of Efforts to Broaden Participation in STEM, Arlington, VA.

Yin, Robert K. (2009). Case Study Research: Design and Methods (4th edition). Thousand Oaks, CA: Sage.

CHAPTER 2: THE POLICY CONTEXT FOR NSF'S PROGRAMS FOR BROADENING PARTICIPATION

Dr. Norman L. Fortenberry
Director, Center for the Advancement of Scholarship on Engineering Education
National Academy of Engineering

PREAMBLE

Among the core values enunciated in the strategic plan (NSF 06-48, 2006) of the National Science Foundation (NSF) is being:

> **Broadly Inclusive**: seeking and accommodating contributions from all sources while reaching out especially to groups that have been underrepresented; serving scientists, engineers, educators, students, and the public across the nation; and exploring every opportunity for partnerships, both nationally and internationally.

Within this context, among the components for consideration as part of evaluation of NSF's merit review criterion on Broader Impacts (NSF Merit Review Board, Web page, 27 Mar 2008) is "How well does the proposed activity broaden the participation of underrepresented groups (e.g., gender, ethnicity, disability, geographic)?" Various advisory and oversight bodies have encouraged NSF's attention to this aspect of the *Broader Impacts Criterion* (CEOSE 04-02, 2004; NSF 04-41, 2004; NSB 04-72, 2004).

However, before evaluators and principal investigators can devote adequate attention to how well they are meeting NSF's goals, they must understand the broader policy context that shaped the development of this core value.

In this brief chapter I summarize the federal legislative, executive, and judicial policy contexts, as well as the impetus provided by various internal constituencies within NSF. In developing this brief chapter, choices had to be made to meet the constraints of available space and time. This is not a comprehensive treatment of the policy context. It explicitly focuses on activities within the three branches of the federal government, and even then focuses primarily on those activities with a direct impact on NSF and its broadening participation programs. For example, there is no discussion of the significant impact of the minority-focused programs at the National Institutes of Health (NIH). This chapter focuses on those activities with tangible and explicit evidence of occurrence and does not address the many cross-currents and couplings that likely existed below the surface and that resulted from the partisan and other political trade-offs that characterize most events in the federal sector. This brief chapter also does not address the critical role played by many nongovernment organizations, most prominently the American Association for the Advancement of Science (AAAS), in providing the data, the discussion venues, and political consensus that influenced many of the developments discussed herein.

This section discusses specific pieces of legislation relevant to broadening participation. However, it should be noted that the context for much legislative action as well as efforts within agencies and the community at large were spurred and informed by numerous education reports issued by the Congressional Office of Technology Assessment (OTA). Many of these reports remain available in an archive maintained by the Federation of American Scientists (FAS Web page, 3 Aug 2008; see “Education” under “E-Topics”).

Many, if not most, of the current programs (or their precursors) directed at broadening participation with respect to underrepresented groups (women, minorities, and persons with disabilities) and institutions at NSF are the result of directive congressional language. Below is a sampling of some of the legislative history.

Women, Minorities, and Persons with Disabilities

Congressional language inserted into the NSF organic act (42 USC Chap. 16: NSF) compels attention by the Foundation to women, minorities, and the economically disadvantaged:

> (a) The Congress finds that it is in the national interest to promote the full use of human resources in science and engineering and to insure the full development and use of the scientific and engineering talents and skills of men and women, equally, of all ethnic, racial, and economic backgrounds.
>
> (b) The Congress declares it is the policy of the United States to encourage men and women, equally, of all ethnic, racial, and economic backgrounds to acquire skills in science, engineering, and mathematics, to have equal opportunity in education, training, and employment in scientific and engineering fields, and thereby to promote scientific and engineering literacy and the full use of the human resources of the Nation in science and engineering. To this end, the Congress declares that the highest quality science and engineering over the long term requires substantial support, from currently available research and educational funds, for increased participation in science and engineering by women and minorities. The Congress further declares that the impact on women and minorities which is produced by advances in science and engineering must be included as essential factors in national and international science, engineering, and economic policies.

The insertion was part of the Science and Engineering Equal Opportunity Act of 1980 (Pub. L. 96-516) which sought to increase participation of women and minorities in science and engineering, and authorizes a wide range of programmatic, evaluative, and oversight activities (including the creation of the Committee on Equal Opportunities in Science and Engineering [CEOSE]) in support of this aim. Later amendments explicitly included persons with disabilities among the populations targeted for participation and advancement in the science and engineering research and education enterprise at all levels.

A crucial aspect of the insertions was the creation of reporting requirements that allow Congress to monitor the Foundation's progress.

Public Law 99-383 (1987) created a Task Force on Women, Minorities, and the Handicapped in Science and Technology with the purpose of developing a long-range plan for broadening participation in science and engineering. Between fall 1987 and spring 1988, public hearings were held around the nation. The Task Force's final report enunciated ambitious goals in six areas: Changing America, PreK-12 Education, Higher Education, Federal Research and Development, Employment, and Influence of Culture. In each area the Task Force made clear that important national goals could not be met without the full participation of women, minorities, and persons with disabilities.

A decade later, the Task Force was echoed by the Congressional establishment in 1998 of the Commission on the Advancement of Women and Minorities in Science, Engineering, and Technology (CAWMSET, or the Morella Commission) (Pub. L. 105-255). This commission had as its mandate to research and recommend ways to improve the recruitment, retention, and representation of women, underrepresented minorities (namely, African Americans, Hispanic Americans, and American Indians), and persons with disabilities in science, engineering, and technology education and employment. The Commission's September 2000 final report (CAWMSET 04-09, 2000) offered recommendations in five areas: Precollege Education, Access to Higher Education, Professional Life, Public Image, and Nationwide Accountability. In each area, the Commission spoke to specific actions that could be taken to enhance the participation, at all levels, in education, research, and practice by women, minorities, and persons with disabilities.

Institutions

The Congress has been equally clear on the need to reach out to institutions beyond the top 50 who historically received the largest chunk of NSF funding. The NSF Authorization Act of 1979 (Barnes, NSF Web page, 3 Aug 2008)[1] compels the NSF director to:

- Operate an experimental program to stimulate competitive research in the interest of assisting States that historically have received relatively little Federal research and development funding and have demonstrated a commitment to improve their research and education programs; and
- Report to specified congressional committees, not later than March 1, 1989, on ways to help academic researchers at the postsecondary level to pursue high-quality research having economic potential.

---

[1] The original text to USC 42 USC 182g Section 113a is not available online, but the equivalent text appears in Public Law 100-570 (the NSF Authorization Act of 1987).

From this direction came the Experimental Program to Stimulate Competitive Research (EPSCoR), first at NSF and eventually at all federal agencies making grants in support of science and engineering research. Additional Congressional impetus for NSF to devote attention beyond the top 50 is given by Committee Report 3 of 500 – Senate Report 106-161 of the FY-00 Appropriations bill (while committee language lacks the force of law, agencies typically recognize that they ignore the guidance provided therein at their future budgetary peril) which indicates (1999):

> Accordingly, the Committee has included a provision to create a focal point for support and outreach to institutions that do not normally fall in the top 50 in federal research and development support. This new office, which will include the highly successful Experimental Program to Stimulate Competitive Research (EPSCoR), is to focus on increasing the Foundation's competitive, merit-based support and outreach to these smaller institutions. The Committee expects NSF to build on its current programmatic and outreach efforts to improve the participation of these institutions and states. The Committee expects the Foundation to submit a detailed proposal for the innovation partnership activity as part of the fiscal year 2000 operating plan.

While it could have been used to create an institutionally-based analog to the state-based EPSCOR program, this language resulted in the creation of the Partnerships for Innovation program which promotes collaborations between research-intensive and non-research-intensive universities.

NSF's authorization for 1988 (Pub. L. 100-570, 1998) also earmarked portions of the FY-89 budget for "the development of model curricula tailored for science and mathematics instruction and instruction in technician training programs, in two-year and community colleges." This later instruction must have been deemed insufficient because the Scientific and Advanced Technology Act of 1992 (Pub. L. 102-476, 1992) requires the NSF director "to carry out a national advanced technician training program of awarding competitive grants to accredited associate-degree-granting colleges which can provide competency-based technical training in advanced-technology occupational fields." This law resulted in the creation of the Advanced Technological Education (ATE) program at NSF.

The NSF authorization for 1988 also:

- Resurrected NSF's programs in undergraduate education (eliminated in the early days of the Reagan administration) by directing NSF to "to support undergraduate science and engineering activities in instrumentation and laboratory improvement, undergraduate faculty enhancement, and research opportunities and curriculum development at the undergraduate level, *as well as efforts to encourage the participation of women, minorities, and the disabled in these fields."* [emphasis added]
- Created an Academic Research Facilities Modernization program with a mandate *"that at least 12 percent of the funds appropriated for the program be set aside for institutions of higher education whose enrollment includes a substantial percentage of Black, Hispanic, or Native American students."* [emphasis added]

THE EXECUTIVE BRANCH CONTEXT

The principle record of attention to broadening participation by the executive office of the President of the United States is principally through executive orders related to specific minority serving institutions. These executive orders mandate the efforts by all federal agencies to strengthen the indicated institutions. With the notable exceptions of the Historically Black Colleges and Universities Undergraduate Program (HBCU-UP) and the Tribal Colleges and Universities Program (TCUP), within NSF this has, since the 1980s, mainly translated into tracking the grants made to these institutions. There were programs directed to minority institutions in the 1970s (e.g., The Minority Institutions Science Improvement Program [MISIP] begun under President Richard Nixon in a precursor to the executive orders referenced above [NSF 81-33, 1981]).

The White House Initiative on Historically Black Colleges and Universities began with the signing of Executive Order 12232 by President Jimmy Carter in August 1980 (executive orders can be searched at http://www.archives.gov/federal-register/executive-orders/disposition.html). Every subsequent president has signed a similar executive order.

The White House Initiative on Tribal Colleges and Universities began with the signing of Executive Order 13021 by President Bill Clinton in October 1996. His successor, President George W. Bush, signed a similar executive order.

The White House Initiative on Educational Excellence for Hispanic Americans began with the signing of Executive Order 12729, signed by President George H.W. Bush in February 1994. Every subsequent president has signed a similar executive order.

THE JUDICIAL BRANCH CONTEXT

An array of programs used to exist at NSF in the 1970s and 1980s to promote the inclusion of women and minorities in the science and engineering research enterprise; however, many of these programs were suspended (or at least no new solicitations were issued) in the late 1980s and 1990s out of concern that they were invitations to lawsuits that could greatly restrict the Foundation's ability to meet its statutory requirement for inclusive programs. Now new solicitations are being issued that reflect the current legal landscape. In a gross oversimplification that serves to make a point, it can be said that programs that once targeted individuals based on their demographic characteristics now specify outcomes to be achieved irrespective of the demographics of the participants. This can be seen, for example, by comparing the 1989 solicitation for the Minority Institutions Science Improvement Program (MISIP, NSF Web page, 3 Aug, 2008) to the 2008 solicitation for the Broadening Participation Research Initiation Grants in Engineering (BRIGE) Program (BRIGE, NSF Web page, 3 Aug, 2008).

The legal environment is complex and rapidly evolving. The author does not claim an ability to do justice to it in the space available. Interested readers are referred to two AAAS reports (though neither specifically discusses NSF programs): the 1996 report on *The Effect of the Changing Policy Climate on Science, Mathematics, and Engineering Diversity* (Malcom, 1996), and the October 2004 report on *Standing Our Ground: A Guidebook for STEM Educators in the Post-Michigan Era* (Malcom, 2004).

THE AGENCY CONTEXT

The agency context for NSF's programs to broaden participation are provided by the policy documents of the National Science Board (NSB) and those CEOSE documents that make specific reference to recommended policy actions by NSF.

The NSB's *2020 Vision* (NSB 05-142, 2005) is that the Foundation will, among other things, "tap the talents of all our citizens, particularly those belonging to groups that are underrepresented in the science and research enterprise, and continue to attract foreign students and scientists to the U.S."

In the NSB's report on *Broadening Participation in Science and Engineering Faculty*, they note (NSF 04-41, 2004):

> For decades, the United States has excelled in building and sustaining institutions of higher education that attract science and engineering talent from all over the world. The Nation has done less well in encouraging and developing the mostly untapped potential of underrepresented minorities, women, and persons with disabilities to contribute to STEM research and education. Developing this potential will lead to expanded opportunities for individuals as well as improving national competitiveness and prosperity.

In the NSB's report on *The Science and Engineering Workforce: Realizing America's Potential* (NSB 03-69, 2003), they note, with regard to undergraduate education, that "the Federal Government must direct substantial new support to students and institutions in order to improve success in [science and engineering] S&E study by American undergraduates from all demographic groups."

In the NSB's report on *Science and Technology Policy: Past and Prologue* (NSB 00-87, 2000), in discussing the role of policy, they observe that cultivating an increasingly diverse student body to renew the workforce of a global economy requires quality science education at the K-12 level. Our education system could serve more students far better than it does, especially those in urban and rural areas born into disadvantage. High standards, expectations, and accountability alone cannot rescue schools lacking the resources to support mathematics and science learning to prepare students for the 21st century workforce. This demands well-trained, well-equipped, and well-rewarded teachers.

In the NSB's report on *U.S. Science and Engineering in a Changing World* (NSB 96-22, 1996), in discussing the needs of current and future generations for a well-trained workforce, they recommend:

> National [science and technology] S&T policies must include a component that addresses the role of science and technology in the development of the Nation's human resource base. This must focus on revitalizing K-12 science and mathematics education at system-wide levels, emphasizing partnerships among diverse communities and all sectors of the economy and encompassing the education and training of S&E personnel in the context of excellence in science, mathematics, engineering, and technology for all Americans.

Agencies' research and development funding decisions have an impact on human resource development. Federal S&T policies should require agencies to take these effects into account when making funding decisions. For example, funding constraints may adversely affect the new partnerships among Federal agencies and laboratories, industry, universities, and schools that emphasize science and mathematics standards in expanding system-wide K-12 education reforms. Likewise, funding decisions have an impact on undergraduate and graduate students, and postdoctoral researchers, affecting both the extent of support to their educational programs and the nature of those programs.

Federal S&T policies should promote the use of networking and information technologies, libraries, museums, community colleges, and S&T centers to increase public understanding of science and technology and to assist the workforce in adopting new skills.

As noted above, CEOSE was created by the Science and Engineering Equal Opportunity Act of 1980. CEOSE is charged with advising the NSF on policies and programs to encourage full participation by women, minorities, and persons with disabilities in science, technology, engineering, and mathematics (STEM). This committee consists of 15 members, each serving a term of three years. The members are researchers and scholars from the STEM fields, and constitute a broad and diverse group drawn from academia, professional organizations, government agencies, and industry. In their retrospective report on broadening participation in America's science and engineering workforce (CEOSE 04-02, 2004), they note that:

> The need—indeed, the imperative—to include ALL Americans in bringing the best of creativity and innovation to the entire STEM enterprise is more vital than ever. The ethical imperatives of equity and justice, along with many pragmatic reasons dictate this need. Among them are the reality of changing demographics, the need to include multiple ways and intelligences to produce the best science and technology, and the changing number of foreign STEM professionals entering the United States. Ensuring broad representation in the STEM workforce is therefore critical.

And make the following recommendations:

Accountability. NSF should expand its systematic and objective evaluation to assess, understand, and report the effectiveness and impact of its programs and policies on broadening participation by:

Continuing to obtain, refine, and disaggregate data and factors related to the participation and advancement of persons from underrepresented groups in STEM education and careers.

Working with the STEM community to develop specific goals, timelines, and metrics, and using them to motivate, track, and hold grantee institutions accountable for progress.

Building assessment and outcome reporting related to broadening participation into NSF program design and accountability expectations where appropriate.

Research. NSF should sponsor additional social science research that will advance understanding of the causes and effects of progress in and barriers to broadening participation in STEM at all levels—from learners to leaders. The relevant individual and institutional factors include mentoring, organizational climate, and the structure, culture, and nature of the systems that constitute the STEM enterprise in the United States. Additionally, NSF should ensure that women, underrepresented minorities, and persons with disabilities are included in the planning and implementation of all research areas, especially those identified for its major investments. It should be noted that the area of "human and social dynamics," identified as one of the areas for major investments by NSF, provides an ideal programmatic framework to include research on these aspects of the STEM enterprise.

Policy Levers. NSF should continue to employ and design new policy levers that focus the attention of principal investigators and their institutions on diversity aspects of the broader impacts criterion, on embedding diversity goals in their research, and on designing and implementing sustainable institutional change that helps STEM become more inviting and supportive of women, underrepresented minorities, and persons with disabilities at all levels.

Tribal Colleges. To engage and advance more Native Americans in STEM, NSF should enhance research capacity and research opportunities at Tribal Colleges by, for example, supporting more faculty exchanges and innovative distance-education and research technologies, expanding collaborations with research institutions, and helping Tribal Colleges and their faculty become competitive at proposal writing and aware of grant opportunities.

It is instructive to view the recommendations from this report, as they illustrate the evolution in methods and approaches in NSF's diversity programming—greater emphasis on accountability metrics, increased reliance on a research base to inform programmatic efforts, a reliance on policy levers to effect outcomes across NSF's portfolio of programs, and greater attention to comprehensive approaches that affect institutional infrastructures affecting all students and faculty.

METRICS

The various laws, orders, and reports appear clear in their aims with respect to members of populations underrepresented in STEM. Women, minorities, and persons with disabilities should be fully engaged in the science and engineering enterprise. Not only should they be well represented among the ranks of students, faculty, and workers, but their academic and professional attainments should mirror those of the general population. Furthermore, the human and societal impacts of science and engineering advances on these populations must be considered essential factors in science and engineering policy. For the research and education communities served by NSF, this would appear to imply, minimally, metrics related to the following:

- Absolute and relative (to the general population and to relevant availability pools) rates of participation in STEM research and education activities and professions by students, staff, faculty, and administrators drawn from underrepresented populations;
- Absolute and relative (to the general population) indicators of institutional policies and practices in support of the participation and advancement of members of underrepresented populations (e.g., scholarship/fellowship support to students, start-up funds for new faculty, institutional matching funds for faculty grants, recruitment and employment practices, maternity/paternity leave policies, etc.); and
- Absolute and relative (to the general population) measures of productivity by members of underrepresented populations (e.g., academic performance, time to degree, journal papers written, grant proposals written, graduate students trained, teaching awards, hours spent advising/counseling students, etc.).

With respect to institutions beyond the top 50, the guidance is less explicit, but the intent appears to be equally clear. Institutions outside the top 50, particularly those serving underrepresented populations, should be assisted in their efforts to become more competitive for NSF research and education grant funds. Given NSF's focus on research, individual non-research universities will probably never be dominant, but just as "the long tail" is a source of innovation in business (Anderson, 2004; Anderson, 2006), similar innovation generators may appear in NSF's portfolio by supporting many more individual institutions with relatively few grants to each.

REFERENCES

Anderson, C. (2004, October). The Long Tail. *Wired*, 12(10).Retrieved from http://www.wired.com/wired/archive/12.10/tail.html

Anderson, C. (2006). *The Long Tail: Why the Future of Business Is Selling Less of More*. New York: Hyperion.

Barnes, D. M. (2007, August.). *National Science Foundation Opportunities and Resources*. Retrieved from http://www.msepscor.msstate.edu/pdf/NSF_EPSCoR_information.pdf Notes: See slide 9 on ESPCoR creation

Commission on the Advancement of Women and Minorities in Science, Engineering, and Technology Development Act. Pub. L. 105-255, 112 Stat. 1889. Retrieved from http://frwebgate.access.gpo.gov/cgi-bin/getdoc.cgi?dbname=105_cong_public_laws&docid=f:publ255.105.pdf

Committee on Appropriations. *Report to Accompany S. 1596*. S. Rep. No 106-161, at 127 (1999).

Committee on Equal Opportunities in Science and Engineering (CEOSE) (2004, December). *Broadening Participation in American's Science and Engineering Workforce, The 1994-2003 Decennial and 2004 Biennial Report to Congress: Executive Summary*. Arlington, VA: CEOSE (CEOSE04-02).

Congressional Commission on the Advancement of Women and Minorities in Science, Engineering, and Technology Development. (2000). *Land of Plenty: Diversity as America's Competitive Edge in Science, Engineering, and Technology*. Arlington, VA: CAWMSET (cawmset0409).

Congressional Statement of Findings and Declaration of Policy Respecting Equal Opportunities in Science and Engineering. 42 U.S.C. 1885 (OSCN 2010), Chapter 16 – National Science Foundation. Retrieved from http://www.oscn.net/applications/OCISWeb/DeliverDocument.asp?CiteID=398691

Federation of American Scientists. *Office of Technology Assessment Archive*. Retrieved from http://fas.org/ota/otareports/topic/etopics/.

Malcom, S. M., Chubin, D. E., & Jesse, J. K. (2004, October). *Standing Our Ground: A Guidebook for STEM Educators in the Post-Michigan Era*. Washington, D.C.: American Association for the Advancement of Science.

Malcom, S. M., George, Y. S., & Horne, V. V. (Eds.) (1996). *The Effect of the Changing Policy Climate on Science, Mathematics, and Engineering Diversity.* Washington, D.C.: American Association for the Advancement of Science.

National Science Board. (1996). *U.S. Science and Engineering in a Changing World*. Arlington, VA: National Science Board (NSB 96-22).

National Science Board. (2000). *Science and Technology Policy: Past and Prologue*. Arlington, VA: National Science Board (NSB 00-87).

National Science Board. (2003). *Science and Engineering Workforce: Realizing America's Potential*. Arlington, VA: National Science Board (NSB 03-69).

National Science Board. (2004). *Broadening Participation in Science and Engineering Research and Education: Workshop Proceedings*. Arlington, VA: National Science Board (NSB 04-72).

National Science Board. (2005). *National Science Board 2020 Vision for the NSF*. Arlington, VA: National Science Board (NSB 05-142).

National Science Board. (2006). *Broadening Participation in Science and Engineering Faculty*. Arlington, VA: National Science Board (NSB 04-41).

National Science Foundation. (1981). *Minority Institutions Science Improvement Program (MISIP): A Brief History 1972-1980*. Washington, D.C.: National Science Foundation (NSF 81-33).

National Science Foundation. (1989). *Division of Research Initiation and Improvement Programs*. Retrieved from http://www.nsf.gov/pubs/stis1989/pol_rii5/pol_rii5.txt?org=EHR

National Science Foundation (2006). *Investing in America's Future: Strategic Plan FY 2006-2011*. Arlington, VA: National Science Foundation (NSF 06-48).

National Science Foundation. (2007). *Broadening Participation Research Initiation Grants in Engineering (BRIGE)*. Retrieved from http://www.nsf.gov/pubs/2007/nsf07589/nsf07589.htm

National Science Foundation. (2007, July). *Merit Review Board Impacts Criterion: Representative Activities*. Retrieved from http://www.nsf.gov/pubs/gpg/broaderimpacts.pdf

National Science Foundation Authorization Act for Fiscal Year 1987. Pub. L. 99-383. Retrieved from http://hdl.loc.gov/loc.uscongress/legislation.99hr4184

National Science Foundation Authorization of 1998, Academic Research Facilities Modernization Act of 1998. Pub. L. 100-570. Retrieved from http://hdl.loc.gov/loc.uscongress/legislation.100hr4418

Science and Engineering Equal Opportunities Act. Pub. L. 96-516. Retrieved from http://hdl.loc.gov/loc.uscongress/legislation.96s568

Scientific and Advanced-Technology Act of 1992. Pub. L. 102-476. Retrieved from http://hdl.loc.gov/loc.uscongress/legislation.102s1146

CHAPTER 3: IMPLICATIONS OF THE NSF BROADER IMPACTS STATEMENT FOR BROADENING PARTICIPATION: AN INCLUSIVE STRATEGY

Donna J. Nelson
University of Oklahoma
Diversity in Science Association

Fitzgerald Bramwell
University of Kentucky

INTRODUCTION

Currently, there is no requirement for grantees to include or gather data on participation by underrepresented groups in most NSF programs, except for several programs in the Division of Human Resource Development (HRD). The only consideration given to broadening participation is the one imposed by the broader impacts requirement. The requirements of and the need for a broader impacts statement appears to be poorly understood and subject to varied and often conflicting interpretation. Under the current rubric, to be compliant with the broader impacts statement does not require the mention of broadening the participation of underrepresented groups. This condition can lead to the continued exclusion of these groups over long periods of time.

In this chapter we offer specific suggestions for actions that can be taken by NSF grantees and staff to address this concern. We also take note of the implications for attention to broadening participation for women, minorities, and persons with disabilities within the larger group of underrepresented groups.

HOW CAN THE BROADER IMPACTS STATEMENT BE USED TO STRENGTHEN BROADENING PARTICIPATION EFFORTS?

The NSF *Dear Colleague Letter on Broader Impacts Proposal Requirements*, dated April 7, 2008 and distributed by email on April 11, 2008, calls the community's attention to several sections of all proposals that require the broader impacts criterion to be specifically addressed. The section elaborating on requirements for the Project Description gives the most detailed description of broader impacts:

> Project Description: Further, as also noted in GPG [Grant Proposal Guide] II.C.2.d., the Project Description must describe, as an integral part of the narrative, the

broader impacts resulting from the proposed activities, addressing one or more of the following as appropriate for the project:

- How the project will integrate research and education by advancing discovery and understanding while at the same time promoting teaching, training, and learning;
- Ways in which the proposed activity will broaden the participation of underrepresented groups, (e.g., gender, ethnicity, disability, geographic, etc.);
- How the project will enhance the infrastructure for research and/or education, such as facilities, instrumentation, networks, and partnerships;
- How the results of the project will be disseminated broadly to enhance scientific and technological understanding; and
- Potential benefits of the proposed activity to society at large.

The NSF *Dear Colleague Letter on Broader Impacts Proposal Requirements* states in its final paragraph: "Since reviewers and NSF program staff must address the broader impacts criterion in the review and decision processes, proposers can draw on examples of broader impacts listed in NSF's *Representative Activities*, and at the [American Chemical Society] ACS Broader Impacts Showcase, but are urged to be creative in their approaches and to discuss ideas with their NSF program officer."

These documents mentioned above are of critical importance in communicating acceptable criteria for broader impacts. Both NSF documents, the *Dear Colleague Letter on Broader Impacts Proposal Requirements* and the *Representative Activities*, refer back to the five bullets shown above.

Of the five groups of activities identified by NSF, only one area—broadening participation—specifically mentions gender, ethnicity, and disability. Therefore, NSF has clearly communicated that the NSF broader impacts requirement may be satisfied by addressing any of the five broad categories of activities, only one of which addresses women, underrepresented minorities, and people with disabilities (WMD) among other underrepresented populations. Although it is commonly perceived that the broader impacts requirement addresses broadening participation by WMD, the actual attention devoted to these groups appears diluted and ineffective. In other words, broadening participation is too often viewed as only one of five different types of activities that can be used to satisfy the broader impacts requirement. We urge the NSF to weave broadening participation issues of diversity, equity, and accessibility specifically into each of these five broader impacts criteria.

The ACS Broader Impacts Showcase presents 34 examples to illustrate the five bullets identified by NSF. The importance of these examples is evidenced by their use as models by other NSF divisions and by organizations outside NSF. Moreover, in private meetings, staff in the U.S. House of Representatives Committee on Science have referred to these activities as constituting excellence in broadening participation. However, we note that some of the 34 Broader Impacts posters convey broadening participation concepts, while in some there is no mention of women, minorities, and persons with disabilities.

Thus, we feel that there is a need to showcase underrepresentation in a creative manner while also addressing broader impacts concerns. For example, collaboration with minority serving institutions or community colleges is reflective of the broadening participation agenda of the Foundation. Thus, an opportunity to advance the Foundation's core value of being broadly inclusive can be leveraged by giving direct attention to the broadening participation activities for each subcategory of broader impacts. We suggest that such examples be examined more closely in order to ensure that they can be used to convey broadening participation concepts as well as those of broader impacts.

We offer the following guidance to measuring broader impacts:

Integrating Research and Education

- Develop and implement quantitative measures of science, technology, engineering and mathematics (STEM) activity participation and the types of institutions involved in collaborative research efforts (e.g., the number of STEM baccalaureates that matriculate to and graduate with terminal STEM degrees);
- Develop and implement quantitative measures of STEM activity participation by underrepresented groups; and
- Develop and implement quantitative measures of STEM activity participation through partnerships with minority serving institutions (MSI) and community colleges.

Building Infrastructure

- Develop and implement quantitative measures for improved policies that promote equitable practices;
- Develop and implement quantitative measures to monitor the dissemination of results from enhanced cyberlearning activities that lead to increased learning and participation for underrepresented groups;
- Develop and implement quantitative measures for changes in performance and perspective based on institutional collaborations such as those with MSIs and other underrepresented institutions; and
- Develop and implement quantitative measures for measuring how an inclusive approach to STEM capacity building has impacted other disciplines.

Broadening Participation

- Described in detail below under the section of this report entitled "Specific Attention to Women, Underrepresented Minorities, and Persons with Disabilities."

Dissemination

- Develop and implement quantitative measures for tracking the usage and accessibility of diverse media by underrepresented groups, with special emphasis on electronic systems, for promoting scientific understanding.

Potential Societal Benefits

- Develop and implement quantitative measures for tracking the application of research and education results by various underrepresented communities, as well as qualitative indicators making a difference.
- Develop and implement quantitative measures for tracking how majority and minority serving institutions provide informal support to one another through dissemination and adaptation.

SPECIFIC ATTENTION TO WOMEN, UNDERREPRESENTED MINORITIES, AND PERSONS WITH DISABILITIES

Foundation-wide templates should require the principal investigator (PI) to formally address broadening participation within the broader impact statement. Examples of areas where reference to broadening participation can be made within these templates include:

Solicitation Criteria

Solicitations can encourage that relevant aspects of the institutional mission statement to broadening participation be included in the discussion of institutional capability.

Recommendation for an NSF Award, Grant, Contract, or Cooperative Agreement

Recommendation documents should reflect the intent to provide quantifiable and measurable evidence of involvement with broadening participation of underrepresented and underserved groups. Appropriate evidence that might be provided includes numbers, percentages, and the distribution of the numbers of individuals involved in and affected by proposed broadening participation efforts.

Annual Progress Reports/Final Reports

There is a need for reliable, consistent, and more detailed data from PIs about students, postdoctoral researchers, and staff supported by their grants. Revisions to data collection methods are needed. For example, existing program-specific data collection efforts need to be coordinated within NSF and linked to the PI annual reports. These reports should use a common set of questions, either across all programs or across program types (e.g., individual investigator research, traineeships and student development grants, and course and curriculum development grants) regarding the number and demographics of:

- Students/postdoctoral researchers recruited for working in the research group;
- Students/postdoctoral researchers joining and working in the research group;
- Students/postdoctoral researchers departing the research group;
- The circumstances surrounding student/postdoctoral researcher departure; and
- The extent of mentoring between the PIs and the students/postdoctoral researchers supported by the grant.

Because the common set of questions suggested above may not enable the PI to provide all the pertinent data, the PI should have the option to provide additional information. Annual reports need a common quantitative reporting system that will yield improved assessment, better data, better highlights, and the advantage of fuller understanding of implementation efforts. What is needed is not just data, but also contextual meaning. The system needs to allow and accommodate longitudinal data.

Program Management Information Systems Reports

Two key questions related to data quality and utility that should be addressed within a quantitative reporting system are:

- How can we get consistent and reliable data?
- How should usable and confidential (across disciplines, racial groups, and gender) data be disaggregated?

Additionally, to be cost-effective, the system needs to allow and accommodate longitudinal data.

NSF principal investigators should:

- Be required to report detailed assessment data annually.
- Track students and postdoctoral researchers, in order to reveal retention and attrition of individuals.
- Follow common terminology established by NSF.
- Have no missing data and no mutually contradictory items.
- Align budget with data collection promises made in the grant proposal.
- Disaggregate data (race, gender, disabilities, national origin).

NSF program officers can facilitate high-quality data collection by:

- Establishing common assessment terminology and communicating this terminology to PIs and all other interested parties.
- Requiring quality control in the information submitted by PIs to NSF about their programs (e.g., no missing data).
- Ensuring that the proposed budget includes funds for appropriate data collection activities.
- Ensuring that promised data collection activities are carried out and reported.
- Verifying that pertinent assessment data were received from each PI in order for that PI and the reporting institution to qualify for future funding.

Instructions to the Committee of Visitors and Advisory Boards

Program officers should also provide appropriate metrics to guide assessment of the broadening participation function at the program/directorate/Foundation level. Such measures that are useful for Committees of Visitors and Advisory Boards might include:

- The percentage of funded grants that embrace the full spectrum of broadening participation activities, those that express some range of broadening participation within the context of broader impacts, and those that make no mention at all of broadening participation;
- The average size and complexity of the program portfolios with broadening participation features managed by individual program officers within the directorate divisions;
- Numbers, percentages, and distribution of numbers of individuals involved in and affected by broadening participation efforts; and
- Proposed and delivered funding of broadening participation efforts.

For grantee-level external advisory boards, additional metrics would include quantifiable measures of:

- Student success in progress toward the degree;
- The percentage of faculty and staff engaged in broadening participation;
- A correlation of grantee expenditures with proposal goal areas; and
- The number, monetary size, and attention to broadening participation in subawards made by the project grant.

Guidance for Review Panels

Frequently, review panelists need guidance for documenting even proposed efforts for broadening participation when the proposal is not submitted to a focused broadening participation program. Because review panels vary in size, composition, and purpose, it may prove difficult to provide generic and useful guidelines. However, we believe that certain underlying themes should form the foundation for their operations. Therefore, we offer the following as a useful but not comprehensive set of examples.

The NSF program officer as well as the principal investigator should consider, as a condition for proposal submission and review or as a condition of acceptance of an award, quantifiable and measurable involvement with broadening participation as reflected in the:

Budget. Funds should be budgeted to promote activities that directly affect broadening participation such as the implementation of mentoring systems, the implementation of evaluation plans, the sharing of useful information through dissemination activities, and the development of useful and effective collaborations for promoting equity and diversity.

Evaluation plans. Such plans should reflect quantifiable measures of broadening participation such as disaggregated baseline measures of student, faculty, and staff engagement; proposed formative and summative measures of tracking and assessing mentoring; matriculation to graduate programs; and transition to the STEM workforce.

Monitoring. Examples of monitoring could include quantifiable measures to be disaggregated by subgroups such as student progress to degree, junior faculty career development, and matriculation from collaborative four-year colleges to graduate institutions, matriculation of STEM majors into the scientific and technological workforce, suggesting the need to request supplemental funding for tracking participants.

Collaborations. Examples of quantifiable measures of effective collaboration might include quantifiable measures of intellectual productivity from diverse perspectives resulting in peer reviewed papers, joint grant submissions, presentations, abstracts, and book chapters.

Dissemination. Dissemination activities should yield effective and quantifiable measures of increased communication throughout an inclusive collaboration, to and within professional organizations, to student affiliates, to relevant participating communities, to local area networks, and to linked web sites.

Management plans. The selection and composition of advisory boards should be in accordance and compliance with institutional guidelines, with attention to representation from populations underrepresented in STEM.

Research plans. Such plans should have some focus on human resource development of underrepresented populations. Where appropriate, this focus should include quantifiable measures of research training and mentoring of junior faculty and students.

Institutional impact (reference to the institutional strategic plan or mission statement). Quantifiable and qualitative measures of institutional impact might include the potential impact of a broadening participation focus/emphasis upon institutional teaching, research, and outreach missions; alignment with institutional strategic goals; and institutionalization of positive outcomes.

For example, if a grant calls for collaboration with a minority serving institution as a means of broadening participation, appropriate funding should be budgeted for the observed project outcome.

SUMMARY

Current efforts concerning broader impacts need enhancement and reformulation to address shortfalls in the realization of outcomes that promote the broadening participation of underrepresented groups, an NSF core value. We find that NSF could benefit from a common, quantitative reporting system for improved assessment, for more reliable data, better access to best practices, and more comprehensive understanding of program strengths and weaknesses. Such a system will require additional follow-up information from proposers and PIs over a sustained period to accommodate longitudinal data, and thereby give the results of prior NSF support additional meaning and perspective.

We suggest that quantifiable and measurable data to measure broadening participation be collected as a requirement of the broader impacts statement as part of the NSF accountability system from proposal submission to submission of the final report. Additionally, we are concerned that broadening participation is too often viewed as only one of five different types of activities that can be used to satisfy the broader impacts requirement. Quantifiable and qualitative diversity, equity, and accessibility data should be collected on awards in each of the five categories listed below, across baseline data, annual reporting, final reporting, and follow-up reporting:

- Integrating research and education,
- Broadening the participation of underrepresented groups,
- Enhancing the infrastructure for research and/or education,
- Disseminating project results broadly to enhance understanding of science and technology; and
- Describing potential benefits to society at large.

Implementation of such an important effort will require the collaboration of NSF administrators, university faculty and staff, and external awardees. These changes should enable documentation of broadening participation within broader impacts.

REFERENCES

Clemson University. (2005). The American Chemical Society (ACS) Broader Impacts Posters. Retrieved from http://chemistry.clemson.edu/NSF-broaderimpactsposters/

National Science Foundation. (2005). Broader Impacts Showcase – ACS Fall 2005 National Meeting & Exposition. Washington, D.C. http://www.nsf.gov/pubs/2005/nsf0540/nsf0540.jsp

National Science Foundation. (2007, July). *Merit Review Board Impacts Criterion: Representative Activities*. Retrieved from http://www.nsf.gov/pubs/gpg/broaderimpacts.pdf

National Science Foundation (2008, April). NSF Dear Colleague Letter on Broader Impacts Proposal Requirements. Retrieved from http://www.nsf.gov/pubs/2008/nsf08044/nsf08044.jsp

U.S. Department of Education. (2007). *Report of the Academic Competitiveness Council*. Washington, D.C.: Retrieved from http://www.ed.gov/about/inits/ed/competitiveness/acc-mathscience/index.html

CHATPER 4: MEASURING SUCCESS AND EFFECTIVENESS OF NSF'S BROADENING PARTICIPATION PROGRAMS: SUGGESTED MONITORING METRICS AND EVALUATION INDICATORS

Beatriz Chu Clewell, Ph.D.
Director, Program for Evaluation and Equity Research
The Urban Institute

INTRODUCTION

In the preceding chapters the authors give the "broader policy context that shaped the development of [the] core value" of being broadly inclusive that is one of the pillars of National Science Foundation's (NSF) *Strategic Plan* (NSF 06-48, 2006). It is clear from these chapters that there is ample validation and justification for NSF's investment in broadening participation across directorates and programs so that underrepresented groups—women, minorities, and persons with disabilities—might fully participate in the scientific and engineering workforce. In addition, the impacts of scientific endeavors on these populations must be considered as important factors in developing science and engineering (S&E) policy. The continued health of the sciences depends on tapping all potential sources of talent. And, as an issue of fairness, all citizens, regardless of race/ethnicity, gender, or disability status, should have the opportunity to participate in scientific professions. Thus, the professional and academic achievements of these groups should be similar to those of the general population.

NSF has discharged its broadening participation responsibilities in the past through a variety of efforts, with varying levels of success. Most recently, in response to the NSF *Strategic Plan*, the Foundation developed a framework for an NSF-wide broadening participation strategy for addressing this key goal through funding and other programmatic efforts. As the strategy is implemented, how can the Foundation measure the success and effectiveness of its investment, especially its investment in programs? This chapter describes the typology of programs funded by NSF that address the broadening participation goal either directly or tangentially and suggests appropriate metrics and indicators that might be used for monitoring and evaluating their effectiveness.[1] Guidance for monitoring and evaluating programs has been provided by the Office of Management and Budget (OMB)[2] and the Department of Education's Academic Competitiveness Council

---

[1] Unless otherwise noted, all programs cited are in the Directorate of Education and Human Resources (EHR) of NSF.

[2] The full text of the OMB Guide to the Program Assessment Rating Tool (PART) can be found at http://www.whitehouse.gov/omb/part/fy2007/2007_guidance_final.pdf

(ACC) report (2007).[3] While such information is agency-oriented, it is also important for awardees to understand Federal reporting requirements.

NSF'S STRATEGY FOR BROADENING PARTICIPATION[4]

The NSF strategy is reflected in the establishment of a funding portfolio that focuses on broadening participation as well as the convening of an NSF-wide Broadening Participation Working Group charged with developing plans to increase participation of underrepresented groups in NSF programs and broadening the pool of reviewers for NSF proposals. The report of the working group titled, *A Framework for Action*, is available on the NSF web site (http://www.nsf.gov/od/broadeningparticipation/nsf_frameworkforaction_0808.pdf). The report, dated August 2008, includes NSF's broadening participation portfolio in Appendix V. (An updated version of NSF's broadening participation portfolio can be found at http://www.nsf.gov/od/broadeningparticipation/bp_portfolio_dynamic.jsp.)

This group, in examining the NSF broadening participation portfolio, categorized the programs as (a) Broadening Participation Focused Programs; (b) Programs with Emphasis on Broadening Participation; (c) Programs with Broadening Participation Potential; and (d) Other Broadening Participation Efforts. The following are descriptions of each program category.

Broadening Participation Focused Programs (28 Programs)

The Broadening Participation Focused programs have the explicit goal of broadening participation and the major part of the budget is dedicated to broadening participation activities (or research on the topic). Within this program category, two major strategies are discernible: (a) broadening the access and success of individuals from underrepresented groups at all levels of the pipeline through provision of resources such as scholarships, fellowships, awards, and interventions; and (b) transforming institutional infrastructure (usually at the postsecondary level) in order to provide learning or work environments that encourage access and success of underrepresented groups in science, technology, engineering and mathematics (STEM).

---

[3] U.S. Department of Education, *Report of the Academic Competitiveness Council*, Washington, D.C., 2007

[4] In addition to individuals who are underrepresented in STEM fields, institutions and geographic regions underrepresented among NSF programs are the focus of NSF's broadening participation efforts.

Included in this category are Broadening Participation in Computing (BPC),[5] the Louis Stokes Alliance for Minority Participation (LSAMP), Increasing the Participation and Advancement of Women in Academic Science and Engineering Careers (ADVANCE), Historically Black Colleges and Universities Undergraduate Program (HBCU-UP), Research in Disabilities Education (RDE), and several other programs. Eighteen of these programs require project-level evaluation in the solicitation.

Programs with Emphasis on Broadening Participation (17 Programs)

Programs with Emphasis on Broadening Participation have an additional review requirement for broadening participation. Awards within the portfolio have broadening participation components as well as other components.

Included in this category are programs such as the Integrative Graduate Education and Research Traineeship Program (IGERT), Research Experiences for Undergraduates (REU), Informal Science Education (ISE), Computer & Information Science & Engineering (CISE) Pathways to Revitalized Undergraduate Computing Education (C-PATH),[6] and several others. Eight of the programs in this category require project-level evaluation in the solicitation.

Programs with Broadening Participation Potential (16 Programs)

Programs with Broadening Participation Potential are programs that have a high likelihood (because of an eligibility criterion or design element) of contributing to broadening participation. Included in this category are Graduate Research Fellowships (GRF), Advanced Technological Education (ATE), STEM Talent Expansion Program (STEP) and Graduate Teaching Fellows in K-12 Education (GK-12). Nine of these programs require project-level evaluation in the solicitation.

Other Broadening Participation Efforts (Five Programs)

Some of these augment core programs to address broadening participation challenges that have been identified within disciplines. Others support innovative experiments in education, interdisciplinary research areas, or partnerships and projects.

They have no discrete solicitation or Dear Colleague Letter.

---

[5] In the Directorate for Computer & Information Science & Engineering (CISE).

[6] In the Directorate for Computer & Information Science & Engineering (CISE).

Examples of other broadening participation efforts include Next Generation Workforce (SBE-like AGEP),[7] Research Partnerships for Diversity,[8] and Significant Opportunities in Atmospheric Research and Science (SOARS).[9]

METRICS FOR MONITORING AND INDICATORS FOR EVALUATION

This section will give examples of relevant metrics and indicators for the first category of programs, the Broadening Participation Focused programs. The examples may also apply to programs in the other broadening participation categories. Indeed, there is little reason why the same monitoring and evaluation measures should not be applied to a broad array of NSF programs with the understanding that expectations of progress will be higher for programs specifically charged with broadening participation. In identifying typical goals of Broadening Participation Focused programs, the author reviewed program descriptions provided by NSF. The goals used in the examples, therefore, are based on actual program goals for the various projects in this category.

Generating Metrics for Monitoring Purposes

Figure 4.1 shows six typical goals of Broadening Participation Focused programs and provides metrics that might guide the collection of both baseline and follow-up data for monitoring purposes. Note that programs may have more than one goal, in which case they may wish to combine metrics from various goals. Note also that the goals span the STEM pipeline from high school through the S&E workforce, reflecting the NSF strategy of increasing the ranks of members of underrepresented groups at every juncture in the pipeline.

There is a need for different types of information for different types of programs. For example, programs focused on human capital in terms of workforce production may consider metrics such as those delineated in Section A of Figure 4.1. On the other hand, programs focused on institutional capacity building would consider monitoring metrics suggested in Section B of Figure 4.1.

---

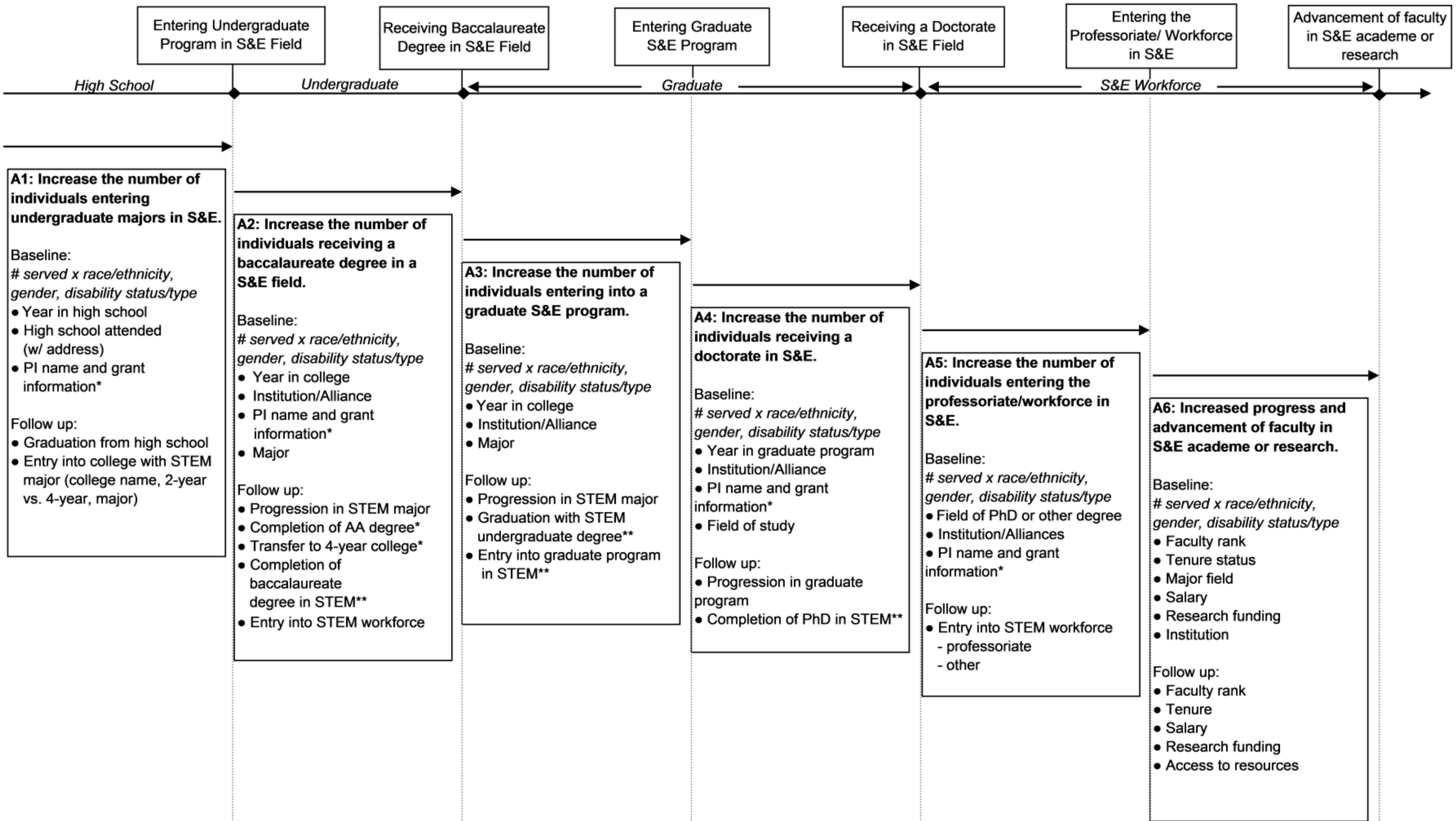
[7] In the Directorate for Social, Behavioral, and Economic Sciences (SBE).

[8] In the Directorate for Mathematical and Physical Sciences (MPS).

[9] In the Directorate for Geosciences (GEO).

Figure 4.1: **Relevant Monitoring Metrics for Broadening Participation Focused Program**

A. Individual-Focused Programs to Increase Diversity of Science & Engineering (S&E) Workforce



** If applicable*

*** Or specialized major field*

Figure 4.1: **Relevant Monitoring Metrics for Broadening Participation Focused Program**

B. Institution-Focused Programs to Increase the Diversity of the Science and Engineering (S&E) Workforce

Goals

B1: Encourage equitable institutional policies and practices in postsecondary STEM departments.

Baseline:
- Distribution of workload (by gender, race, etc.)
- Distribution of institutional resources (space, salaries, etc., by gender, race, etc.)
- Distribution of research funding (by gender, race, etc.)
- Breakdown of STEM faculty (by tenure status, rank, gender, race, etc.)
- Institution type

Follow up:
- Distribution of workload (by gender, race, etc.)
- Distribution of institutional resources (space, salaries, etc. , by gender, race, etc.)
- Distribution of research funding (by gender, race, etc.)
- Breakdown of STEM faculty (by tenure status, rank, gender, race, etc.)

B2: Increase research capability and teaching effectiveness in S&E disciplines.

Baseline:
- Collaborative relationships
- Funding level distribution
- % of URM enrollment
- Total enrollment
- MSI status
- Institution type (Carnegie Classification)

Follow up:
- Collaborative relationships established
- Funding support obtained
- Teaching reform effected

B3: Encourage collaboration of MSIs w/ other entities to enhance effectiveness.

Baseline:
- Collaborative relationships existing
- Research funding levels
- MSI status (type)
- Institution type (Carnegie Classification)
- Enrollment

Follow up:
- Collaborative relationships established
- Research funding support obtained

Monitoring data are used to determine at short-term intervals (typically, one year or less) whether programs are on target in meeting established benchmarks. After collecting baseline data against which to measure progress, follow-up data may be collected on a yearly basis for the period of the intervention or even beyond.

The metrics shown in Figure 4.1 are based on the broad goal categories represented in the set of Broadening Participation Focused programs reviewed. Depending on the specific individual programs being monitored, program administrators or evaluators may wish to add, delete, or revise some of the metrics given.

Developing Indicators for Evaluation

Indicators, developed to answer evaluation questions, differ somewhat from metrics that are used for monitoring purposes. It is certainly desirable that monitoring data be used, if possible, for evaluation purposes. For this to happen, however, it is especially important that the data collected for monitoring purposes have validity.[10] This requires that validation checks be made as data are being collected and that response rates be sufficient to ensure that the data are representative of the population.[11] It is also necessary to augment monitoring data with other data collected specifically for evaluation purposes such as, for example, a retrospective survey of program participants to collect follow-up data not included in the monitoring data.

Figure 4.2 shows possible indicators of success for Broadening Participation Focused Programs developed in response to an evaluation question addressing a program goal. The goals, questions, and indicators are divided by the two main target foci for Broadening Participation Focused Programs: individuals (workforce diversification) and institutions (institutional transformation in the interest of workforce diversification).

Again, as with the examples of monitoring metrics, these examples may also apply to other categories of broadening participation programs.

[10] Although data used for monitoring should also have validity, this has not always been the case.

[11] For a discussion of data quality, including data validation and response rates, please see Campbell and Clewell (2008).

Figure 4.2: Evaluation Goals, Questions, and Related Indicators: Broadening Participation Focused Programs

A. Individual-Focused Programs to Increase Diversity of Science & Engineering (S&E) Workforce

1. Increase the number of individuals entering into an undergraduate program in S&E.

Question	Indicators
Has the program increased the number of targeted individuals entering into an undergraduate program in S&E (or, where applicable, a specific major) ?	• # and % of participants (compared to a comparison sample) who enroll in a postsecondary institution with the intention of majoring in S&E (or, where applicable, a specific major) • Distribution of participants who entered 2-year versus 4-year colleges • Distribution of intended majors among participants • Racial/ethnic and gender breakdown of successful participants • Graduating GPA of participants (average, successful vs. unsuccessful, etc.)

2. Increase the number of individuals receiving a baccalaureate degree in a S&E field.

Question	Indicators
Has the program increased the number of targeted individuals receiving a baccalaureate degree in an S&E field?	• # and % of participants (compared to a comparison sample) who complete a baccalaureate degree in S&E (or specific major) field • # and % of participants who enrolled in 2-year colleges who transfer to a 4-year college with a S&E major • Racial/ethnic and gender breakdown of participants who completed baccalaureate degrees • Graduate GPA of participants (average, successful vs. unsuccessful, etc.)

3. Increase the number of individuals entering into a graduate S&E program.

Question	Indicators
Has the program increased the number of targeted individuals entering into a graduate program in S&E?	• # and % of participants (compared to comparison sample) who enter a postgraduate program in S&E (or specific major) field • # and % of participants who enter master's programs vs. Ph.D. programs • Racial/ethnic and gender breakdown of participants who enter graduate programs • Distribution of successful participants among S&E fields • Distribution of successful participants among institutions by type (Carnegie Classification)

4. Increase the number of individuals receiving a doctorate in S&E.

Question	Indicators
Has the program increased the number of targeted individuals completing a doctorate in S&E?	• # and % of participants (compared to comparison sample) who complete a doctorate in S&E (or specific major field) • Racial/ethnic and gender breakdown of successful participants vs. unsuccessful participants • Distribution of successful participants among S&E fields • Distribution of successful participants by institution type

5. Increase the number of individuals entering the professoriate/workforce in S&E.

Question	Indicators
Has the program increased the number of targeted individuals entering the professoriate/workforce in S&E?	• # and % of participants (compared to comparison sample) who are employed as faculty in S&E departments • Racial/ethnic and gender breakdown of successful participants vs. unsuccessful participants • Breakdown of institution type where successful participants are employed as faculty • Distribution of successful participants among faculty ranks • # and % of participants (compared to comparison sample) who are employed in the S&E workforce (outside of academe) • Racial/ethnic and gender breakdown of successful participants vs. unsuccessful participants

6. Increase progress and advancement of faculty in S&E academe or research.

Question	Indicators
Has the program contributed to the increased progress and advancement of S& E faculty from targeted groups	• # and % of participants (compared to comparison sample) who are promoted, receive tenure, achieve salary equity and receive research funding as well as other types of resources • Racial/ethnic and gender breakdown of participants • Distribution of participants among faculty ranks • Distribution of participants among institutional types

Figure 4.2: Evaluation Goals, Questions, and Related Indicators: Broadening Participation Focused Programs

B. Institution-Focused Targeted Programs to Increased Equity, Productivity, and Efficacy of Institutions (Including Minority Serving Institutions [MSI]) in Science and Engineering (S&E)

1. *Encourage equitable institutional policies and practices.*

Question	Indicators
Has the program resulted in the adoption of institutional policies and practices to increase equity?	• Establishment of equity-friendly institutional policies and practices or changes in institutional policies and practices to encourage equity

2. *Increase research capability and teaching effectiveness in S&E disciplines.*

Question	Indicators
Has the program resulted in increased research capability in S&E departments?	• Increases in reported research funding received
Has the program resulted in increased teaching effectiveness in S&E courses?	• Evidence of teaching reform (revised course syllabi, new courses established, workshops held on advanced teaching techniques, etc.)

3. *Encourage collaboration of MSIs w/ other entities to enhance effectiveness.*

Question	Indicators
Has the program resulted in increased collaboration of MSIs with other entities to increase research teaching effectiveness?	• Evidence of increase in collaborations • Increases in research funding levels

Please note that the indicators' lists contain frequent references to comparison samples. These samples may be randomly selected samples, matched samples, or samples from national data bases, as described in the following chapter. Where increases in individuals or institutional capabilities are being measured, rates of increases or absolute numbers may be compared to those of a comparison sample. See Chapter 6 (p. 64) which provides research designs.

USES OF EVALUATION DATA FROM BROADENING PARTICIPATION PROGRAMS

How evaluation data from broadening participation programs are used will vary with the type of user. Typically, users of these data are policymakers, funders, principal investigators, and evaluators. Data derived from monitoring metrics usually serve different purposes than do data reported from evaluations. The following section describes ways in which monitoring metrics and evaluation data might be used by different groups.

Policymakers

Policymakers are most likely to use evaluation data from summative evaluations rather than monitoring metrics. Summative evaluations report on the ability of programs (as opposed to projects) in a given agency to meet their stated goals.[12] The results of these evaluations may help Congress to determine how much funding to allot to NSF and how these funds should be distributed. In the specific case of the broadening participation programs at NSF, Congress and federal agencies will be able to determine how well the NSF is fulfilling its goal of increasing the diversity of the S&E talent pool and workforce, reward it accordingly, and/or suggest changes in the strategies that are being employed to attain this goal.

The National Science Foundation

Monitoring metrics also provide useful information within the division or program area to gauge the progress of various programs in meeting established short-term benchmarks. These metrics might be reported to NSF by programmatic areas or divisions in order to ensure continued funding when summative cross-project evaluation data are not yet available.

---

[12] Data from monitoring metrics, on the other hand, can be used as a short-term gauge of program progress towards meeting goals, influencing decisions as to continuation of funding.

Individual Projects

Principal investigators, program staff, and evaluators in individual projects rely on data from monitoring metrics to assess the progress of projects as they are being implemented. By comparing monitoring metrics against pre-established benchmarks, project administrators may pinpoint problems at early implementation stages as well as justify to NSF that the project is being carried out in a timely manner. These groups also may use evaluation data to report outcomes to NSF, their advisory boards, and host institutions. Evaluation data may, furthermore, be reported to NSF at the completion of a project to show the extent to which the project met stated goals. If the evaluation data show that the project was successful, principal investigators might use the data to generate additional funding from either NSF or another funder, or to lobby for institutionalization by the host institution.

The Research and Practitioner Community

Evaluation data (but probably not data from monitoring metrics) are useful to the research community in expanding the knowledge base regarding interventions to broaden participation. Practitioners who run interventions to broaden participation can also learn from successful (or even unsuccessful) broadening participation efforts and may wish to replicate practices that have been found to be especially effective.

NSF'S ROLE IN TRACKING THE OUTCOMES OF BROADENING PARTICIPATION EFFORTS

The *Broadening Participation* report of the Working Group calls for NSF to "conduct periodic evaluations, including external reviews ranging from the program level to larger cross-sections of the portfolio" (2008, p. 14). In order to accomplish this, program areas at NSF should develop a common framework requiring that all broadening participation projects collect uniform data via both monitoring metrics and evaluation indicators. This will facilitate the cross-project and cross-program review of broadening participating efforts and allow NSF to conduct reviews of larger cross-sections of the broadening participation portfolio as called for in the report.[13] The data collection framework should be described in the RFPs for these projects and respondents to the RFPs should be asked to include data collection according to this framework in their evaluation plans. The Working Group report also calls for evidence of broadening participation activities to be reported in existing mechanisms such as "performance plans, divisional reports, and external evaluations" (2008, p. 15). As part of the process of tracking outcomes of broadening participation initiatives funded by NSF, it is important that longitudinal data be collected to determine the longer-term impact of the set of initiatives (for a discussion of the need for longitudinal tracking, please see page 65 of this report).

[13] For a guide to designing and implementing cross-project evaluations, please see Clewell and Campbell (2008) and Campbell and Clewell (2008).

Finally, to ensure a broader application of its broadening participation mandate, NSF should review all funded programs to determine the following:

- *Are program funds serving a representative proportion of members of underrepresented groups or institutions?* Indicators to address this question might include the number and percent of participants served disaggregated by race/ethnicity, sex, and disability status, and number and percent of institutions awarded funding disaggregated by Carnegie classification, minority serving institution (MSI) status, and region of country.

- *Are positive outcomes of programs (as reflected in evaluations) distributed equitably among all groups of participants or institutions, including underrepresented groups and institutions?* Indicators to address this question might include the number and percent of participants showing positive outcomes by race/ethnicity, gender, and disability status, and number and percent of institutions showing positive outcomes by Carnegie classification, MSI status, and region of country.

REFERENCES

Campbell, P.B. & Clewell, B.C. (2008), *Building Evaluation Capacity: Guide II Collecting and Using Data in Cross-Project Evaluations*. Washington, DC: The Urban Institute. Retried from http://www.urban.org/UploadedPDF/411651_guide2.pdf

Clewell, B. C. & Campbell, P. B. (2008, January). *Building Evaluation Capacity: Guide I Designing A Cross Project Evaluation*. Washington, DC: The Urban Institute. Retried from http://www.urban.org/UploadedPDF/411651_guide1.pdf

National Science Foundation. (2008, August). *Broadening Participation at the National Science Foundation: A Framework for Action*. Arlington, VA: NSF. Retrieved from http://www.nsf.gov/od/broadeningparticipation/nsf_frameworkforaction_0808.pdf

National Science Foundation Broadening Participation Portfolio. Retrieved from: http://www.nsf.gov/od/broadeningparticipation/bp_portfolio_dynamic.jsp

U.S. Department of Education. (2007). *Report of the Academic Competitiveness Council*. Washington, D.C.: Retrieved from http://www.ed.gov/about/inits/ed/competitiveness/acc-mathscience/index.html

PART II: FOCUS ON THE IMPACTS OF STEM BROADENING PARTICIPATION PROJECTS

CHAPTER 5: OUTCOMES AND INDICATORS RELATING TO BROADENING PARTICIPATION

Patricia B. Campbell, Ph.D.
Campbell-Kibler Associates, Inc.

Veronica G. Thomas, Ph.D.
Howard University

Adam Stoll, Ph.D.
Congressional Research Service

In most assessment areas, it is useful to begin with definitions. Outcomes "refer to what is ultimately achieved by an activity, as distinct from its outputs which relate to more direct or immediate objectives." (Glossary of Statistical Terms, n.d.). Indicators provide evidence that a certain condition exists or certain results have, or have not, been achieved. They can measure inputs, process, outputs, and outcomes. Horsch (2006) described a number of different types of indicators including:

- Input indicators, which measure resources, both human and financial, devoted to a particular program or intervention and can include measures of characteristics of target populations;
- Process indicators, which measure ways in which program services and goods are provided;
- Output indicators, which measure the quantity of goods and services produced and the efficiency of production; and
- Outcome indicators (or in our case, just outcomes) which measure the broader results achieved through the provision of goods and services. These indicators can exist at various levels: population, agency, and program.

Quantitatively

At a quantitative level, a major outcome of broadening participation is to increase the number and diversity of members from underrepresented groups in each of the following five areas:

- Having access to the benefits of science, technology, engineering, and mathematics (STEM) knowledge,
- Having access to STEM knowledge,
- Studying STEM,
- Working in STEM areas; and
- Generating STEM knowledge.

At first glance, assessing this outcome seems technically difficult, but straightforward. Unfortunately that is not the case. Before this, or any outcome, can be assessed, definitions of success need to be established. In the case of broadening participation, there are a number of challenges related to defining success. Defining success in terms of increases in absolute numbers or percentages is often insufficient at best, and misleading at worst. A 300 percent increase in, for example, Native Americans receiving Ph.D.s in engineering sounds very impressive but it could be merely representing an increase from one to three. Defining success as an increase in both the number and percentage of increase is better but not adequate; in part because it does not set success within a broader context. For example, it should not be seen as success in broadening participation if the percentage of African Americans taking STEM Advanced Placement (AP) courses increases but the percentage of students taking STEM AP courses who are African American decreases (note this would happen if the numbers of other students taking STEM AP course increased at a higher rate than that of African American students). Defining success as “reducing the gaps while all gain” puts change in context but does not provide an end point—a point when success is achieved.

One definition of success that includes an “end point” is parity, which is defined by Merriam Webster (2008) as “the quality or state of being equal or equivalent.” In the cases of the first two areas, having access to the benefits of STEM knowledge and having access to STEM knowledge, parity could be with the population of the United States. That is members of designated subgroups (i.e., underrepresented minorities [URM], women, people from rural areas) have the same access as does the population as a whole. For area three, studying STEM, parity may be the percentage of the subgroup studying STEM that reflects the subgroup’s percentage in the general population. That is, if African Americans are 12.8 percent of the population, then parity in the third area is African Americans being 12.8 percent of those studying STEM. For the fourth area, working in STEM areas, parity may be where the percentage of the subgroup working in STEM reflects the percentage of that group studying STEM. For the fifth area, generating STEM knowledge, parity may be where the percentage of a subgroup generating STEM knowledge reflects the percentage of that group working in STEM.

Realistically, parity should be defined as a range rather than an absolute number. For example, rather than have 12.8 percent be defined as parity for African Americans studying STEM, a range would be used such as 10 to 15 percent. Of course that range is an example; a process and rationale would need to be developed to determine and justify appropriate ranges.[1]

Since parity is about being equal or equivalent, there should be equal concern when a group is over the range as well as under the range. Over participation of a subgroup in a field should be as great a concern as under participation. In the case of gender, the “feminization” of a field has traditionally resulted in a loss of status and income and may contribute to men’s avoidance of fields they perceive as “female” (England, 1992). By the same token, women may shy away from fields perceived as typically male.

While parity is an intriguing definition of success, even it is not sufficient. Success in broadening participation must include increases in absolute numbers as well. Perhaps success can be best defined as achieving parity as more participate overall.

Regardless of what definition of success is used, it will also be necessary to define the size of the fields or discipline areas to which the definition of success is applied. If a field size is too broad then areas in need of having participation broadened can be missed. For example, while women's enrollment in engineering is around 20 percent, there is great variability by field: almost half (43.7 percent) of the biomedical engineering students are women compared to women being only 12.5 percent of the mechanical engineering students (Engineering Workforce Commission, 2008). On the other hand, if a field size is too narrowly defined, there may be privacy and confidentiality issues related to the small number of underrepresented participants in that field. Thus, balance must be achieved in terms of specificity of field and subfields within a particular STEM discipline.

Qualitatively

As important as the numbers are; they should not be the only indicators of success. A stronger, more robust STEM knowledge base is another outcome of broadening participation as the use of diverse perspectives and populations in STEM research and development efforts increase. For example, Fortmann, Ballard, and Sperling (2008) reported that a collaboration between plant breeders (scientists) and bean experts (Rwandan farmers) showed that “real technical gains can be achieved through direct collaboration (and knowledge sharing) with farmer experts, in this case, women. Plant criteria essential to farmers are brought to the fore, production gains are achieved and variety diversity is promoted,” and research cost-effectiveness was improved. Changes to include diverse perspectives— and, where appropriate, diverse populations—in research studies have the potential to be transformative.

[1] While issues related to the presentation of data are beyond the scope of this chapter, it is important to note the value of contextualizing percentages and numbers with trends, progressions, and slopes

In order to develop a comprehensive strategy aimed at demonstrating the success and effectiveness of the NSF's investment in broadening participation, it is essential that indicators of success be considered at three levels—individual level, institutional level, and NSF level. Indicators at each of these three levels might represent (a) *inputs*, or the resources, contributions, and investments made to support broadening participation; (b) *outputs,* or the size and/or scope of the activities, services, events, and products that reach underrepresented groups in broadening participation initiatives; (c) *process*, including the extent to which broadening participation projects, programs, and strategies were delivered as intended; and (d) *outcomes* which are the changes for individuals, groups, communities, institutions, and systems, and may include both short-term or long-term results. Table 5.1 (p. 60) summarizes the broadening participation indicators across the individual, institutional, and NSF level.

Broadening Participation Indicators at the Individual Level

Valid and reliable student level data are essential for determining the causal pathways for success of broadening participation initiatives. Therefore, indicators at the individual level generally focus on myriad student level data. Generally, student level indicators can be identified across four dimensions, including: (a) participation, (b) retention, persistence, and success, (c) experiences, and (d) attitudes.

Participation, in its purest sense, refers to the total share that students from underrepresented groups have of the total student enrollment within a particular STEM field or major. Some key indicators of participation include presence of students from underrepresented groups in STEM courses and STEM majors. The presence of a critical mass of students from underrepresented groups in STEM majors is also another indicator of participation.

Persistence or retention refers to the number of students from underrepresented groups who both return to school and remain in a STEM field. Success generally refers to the proportion of students from underrepresented groups who completed their STEM degree program. Here, indicators include, for example, year-to-year persistence, success at the course level, grade point average (GPA), time to degree, graduation rates, and attrition rates. Another more indirect indicator includes student receipt of honors, awards, and recognition in STEM-related areas.

Student experiences include various indicators designed to both document and understand the experiences of students from underrepresented groups which may contribute to their success (or lack thereof). These indicators include, for example, involvement in research experiences at key points during undergraduate and graduate training, participation in bridge or mentoring programs, mentoring by active researchers, presentations and/or attendance at scientific meetings, and intellectual and social networking with peers (see Table 5.1 for more elaboration).

Attitudes refer to a host of indicators focusing on students' perceptions and attitudes toward STEM fields and toward themselves in relation to STEM interest and confidence.

Sample indicators include confidence in ability to excel in STEM fields and positive attitudes toward STEM discipline and careers.

Broadening Participation Indicators at the Institutional Level

Clearly, broadening participation must go beyond addressing the individual student needs and focus attention on ensuring deep-seated institutional transformation. This involves a change from how institutions have traditionally operated to an explicit emphasis on equity, inclusion, and broadening participation of individuals (as students, faculty, and administrators) from underrepresented groups. Indicators at the institutional level can be examined across five areas: (a) staffing, (b) policies, programs, and institutional commitment, (c) accountability and rewards, (d) monitoring, tracking, and using data for improvement, and (e) collaborations.

Staffing, in this section, refers to institutional personnel, in particular faculty from underrepresented and non-underrepresented groups at colleges and universities. Attention must be given to the collection of data on STEM faculty at research institutions, particularly faculty from underrepresented groups. Indicators include the demographics of existing STEM faculty and the hiring patterns of new faculty from underrepresented and non-underrepresented groups at both research and non-research level institutions.

Policies, programs, and institutional commitment refer to myriad strategies in place at the institution that demonstrate the emphasis and commitment to broadening participation of underrepresented groups. Examples of key indicators include the centrality of diversity and broadening participation in institutional strategic planning, presence of a written diversity mission statement, institutional budget/resources dedicated to diversity and closing gaps, and the presence of institutional plans, programs, and services aimed at broadening participation and enhancing the success of underrepresented populations (see Table 5.1 for elaboration).

Accountability and rewards focus on the presence of accountability metrics in broadening participation initiatives at the institution. Sample indicators include the presence of a strategy for holding institutional leadership (vice presidents/deans) accountable (e.g., in performance appraisals) for facilitating broadening participation efforts, inclusion of broadening participation accomplishments in faculty activities reports and departmental/school/college/institutional quarterly/semi-annual/annual reports, and the expected involvement of senior, nonminority faculty in efforts to broadening participation (see Table 5.1 for elaboration).

Monitoring, tracking, and using data relate to the institution's formal mechanisms for gathering and utilizing data to enhance broadening participation efforts. Indicators include the presence of a framework for monitoring the success of broadening participation efforts (including a tracking system to follow student progression and success) and the use of data for program improvement.

Collaborations include focus on relationships and partnerships with various entities serving large segments of students from underrepresented groups. This includes the presence of two-way partnerships between majority and minority serving institutions (MSI), K-12 schools, communities, and various organizations (e.g., Boys and Girls Clubs).

Broadening Participation Indicators at the NSF Level

Indicators for broadening participation are important at the funder and Foundation level as well. Externally it is the funder who often serves as the catalyst for broadening participation efforts at the institutional level and whose requirements fuel the collection of data at the individual level. The funder also has the responsibility to broaden participation internally as well. The following indicators cover both internal and external areas.

Inclusion of information about the importance of broadening participation. Those submitting proposals to NSF as well as those serving as reviewers, advisors, and as members of Committees of Visitors (COV) look to the materials that are provided to them for guidance on what is important to NSF. Those writing proposals want to increase the probability of funding while others want to provide NSF with information that will be useful. Including information on the importance of broadening participation is an indication that this is important to the Foundation.

Review and monitoring of Foundation policies and practices in terms of their potential to broaden participation. Following the model of the environmental impact statements, policies and practices, even if they are not directly related to broadening participation, should be reviewed for potential impact. For example, a practice of using existing reviewers' nomination of additional reviewers as a major source of new reviewers can have a negative impact on broadening participation.

Diversity of professionals involved with NSF. Numbers and percents of people with different demographic characteristics who are involved with NSF should be documented. This includes internal participants such as program officers, division directors, and others, as well as external participants such as reviewers, advisory panel members, COV members, and principal investigators.

Foundation resources devoted to broadening participation. Resources can include staff allocated to broadening participation efforts, intervention programs, training opportunities, materials, travel money, and project supplements targeting broadening participation.

***Foundation resources devoted to research on broadening participation*.** The number of research projects on broadening participation, the number of research projects including a focus on broadening participation, the number of divisions with any such research studies, and the total amount of money provided to such studies can all be included here.

Improvements to the knowledge base about broadening participation. Improvements may include process variables, such as being able to link across different data sets; requiring oversampling of underrepresented populations in longitudinal and other national and state data collection efforts as increased knowledge about barriers to broadening participation; and about what works and what doesn't in different environments to broaden participation.

Implementation of strategies found to be effective at the Foundation level to broaden participation. This can include dissemination of strategies found to be effective both within NSF and to audiences outside of NSF; monitoring within NSF to see if effective strategies are being implemented; and collecting information as to the degree to which effective strategies are implemented by those receiving NSF funding as well as by other institutions and organizations.

Table 5.1: Indicators for Evaluating Broadening Participation Initiatives.

A. Individual Level Indicators
Participation of Underrepresented Students (URS) • Presence of URS in STEM courses • Presence of URS in STEM majors at undergraduate and graduate levels • Existence of a critical mass (cohort) of URS
Retention, Persistence, and Successful Completion for Underrepresented Students • Year-to-year persistence • Success at the course level • Success at the degree level/graduation rates • GPA • Honors, awards, and recognitions • Time to degree (by demographic factors, STEM discipline, and student experiences) • Attrition rates (when and why; what majors do STEM student change to) • Trends over time
Experiences of Underrepresented Students (Dosage) • Involvement in research experiences midway through undergraduate training (including quality of research experiences) • Involvement in research experiences early during graduate training • Presentations and/or attendance at scientific meetings (locally, regionally, nationally, or internationally) • Participation in bridge or mentoring programs to help encourage and retain (particularly undergraduate students) interest in STEM • Mentoring by active researchers • Advisement • Intellectual and social networking with peers
Attitudes • Confidence in ability to excel in STEM field • Positive attitudes toward STEM discipline and careers • Knowledge of value of STEM in workplace and everyday life

Table 5.1: Indicators for Evaluating Broadening Participation Initiatives.

B. Institutional Level Indicators
Staffing • Demographics of existing STEM faculty (by rank, gender, ethnicity, tenure, and salary) at research and non-research universities • Hiring patterns of new STEM faculty (by rank, gender, ethnicity, tenure, and salary) at research and non-research universities
Policies, Programs, and Institutional Commitment to Broadening Participation • Centrality of diversity and broadening participation in institutional strategic planning • Presence of written and widely articulated diversity mission statement • Presence of designated institutional office (e.g., Office of Institutional Diversity) and officer focusing on broadening efforts, including the real and perceived authority of the office/officer • Institutional budget/resources dedicated to diversity and closing gaps • Presence of institutional plans, programs, and services aimed at broadening participation and enhancing the success of underrepresented populations (e.g., funds to address student needs related to broadening participation, funding of fellowships and postdocs, mentoring programs, bridge programs, tutoring programs, etc.) • Diversity of faculty participating in student recruitment and curriculum transformation
Accountability and Rewards • Presence of mechanisms for holding institutional leadership (e.g., vice presidents/deans) accountable (e.g., in performance appraisal) for how well they facilitate broadening participation efforts and resources to support their departments/units • Inclusion of broadening participation accomplishments in faculty reports and institutional reports at various levels (unit/department/school/college/institution-wide) • Provision of incentive/rewards for faculty development (e.g., use of technology to engage underrepresented groups) and scholarship development related to broadening participation • Expected involvement of senior-level, non-minority faculty in broadening participation efforts
Monitoring, Tracking, and Using Data for Improvement • Presence of a framework for monitoring the success of broadening participation efforts (including a tracking system to follow student progression and success) • Proactively collecting, utilizing, and disseminating data to enhance the success of broadening participation
Collaborations • Presence of two-way partnerships between majority institutions and minority serving institutions • Presence of partnerships with K-12 schools, communities, and organizations serving significant proportions of students from underrepresented groups

Table 5.1: Indicators for Evaluating Broadening Participation Initiatives.

C. Foundational Level Indicators
Inclusion of information about the importance of broadening participation, within and across divisions in: • Grants announcements and requests for proposals • Reviewer materials and orientations • Advisory panel materials and orientations • Committee of Visitor review criteria
Review and monitoring, within and across divisions of policies and practices in terms of their impact on broadening participation
Diversity of professionals involved with NSF, within and across divisions, in the following areas: • Reviewers • Advisory panel members • Committees of Visitors • Principal investigators • Program officers • Foundation administrators.
Foundation resources devoted to broadening participation efforts and assessment of the effectiveness of such efforts
Foundation resources devoted to research on broadening participation
Improvements to the knowledge base about broadening participation
Implementation of strategies found to be effective at the Foundation level to broaden participation

REFERENCES

Engineering Workforce Commission. (2003-2010). *Data*. Retrieved from http://www.ewc-online.org/data/index.asp

England, P. (1992). *Comparable Worth: Theories and Evidence*. New York, NY: Aldine de Gruyter.

Fortmann, L., Ballard, H. & Sperling, L. (2008). Change around the edges: Gender analysis, feminist methods and sciences of terrestrial environments. In Schiebinger, L. (Ed.) *Gendered innovations in science and engineering* (pp 79-96). Stanford, CA: Stanford University Press.

Outcomes. (n.d.). In *Glossary of Statistical Terms*. Retrieved from http://stats.oecd.org/glossary/detail.asp?ID=4796

Horsch, K. (1997). *Indicators: Definition and Use in a Results-Based Accountability System.* (Reaching Results). Retrieved from http://www.gse.harvard.edu/hfrp/pubs/onlinepubs/rrb/indicators.html

Parity. (n.d.). In *Merriam-Webster's online dictionary*. Retrieved from http://www.merriam-webster.com/dictionary/parity

U.S. Census: State and Country Quick Facts (n.d.). In U.S. Census Bureau online data set. Retrieved from http://quickfacts.census.gov/qfd/states/00000.html

CHAPTER 6: EVALUATING EFFORTS TO BROADEN PARTICIPATION IN STEM FIELDS

Patricia B. Campbell, Ph.D.
Campbell-Kibler Associates, Inc.

Adam Stoll, Ph.D.
Congressional Research Service

Veronica G. Thomas, Ph.D.
Howard University

Over the last decade, at virtually all levels of practice, education-related programs are being held more accountable for results than was formerly the case. Most of the large educational reform efforts over that time period at the federal and state level have focused on educational standards and accountability. Generally speaking, there is growing interest in information on the effectiveness of educational programs and the policies that support and craft such programs. A natural outgrowth of this is increased interest in and attention to program evaluation. In the case of broadening participation programs, evaluation can play a critical role in demonstrating (short-term and long-term) success of these efforts. Success can be defined in multiple ways and at various levels in the evaluation. For example, at the individual level, success might be defined in terms of enhancing awareness of science, technology, engineering and mathematics (STEM) fields among individuals from underrepresented groups. Another individual level broadening participation success indicator might be broadening the access to STEM fields among individuals from underrepresented groups, and increasing their participation in such programs. At the institutional level, success might be defined in terms of the presence of institutional plans, programs, and services aimed at broadening participation and enhancing the success of underrepresented populations in STEM fields.

There is clearly growing interest in summative evaluations of broadening participation efforts, which offer the promise of confirming the efficacy of a program approach and revealing whether continued investments are warranted. Perhaps more than in past eras, substantial attention is being paid to methodological rigor, with most funding agencies and evaluators feeling strongly incentivized to employ methodologies that will stand up to scrutiny. While it is always desirable to select the most rigorous methodology available, often the demand for higher levels of evidence must be balanced against an array of practical considerations that affect evaluation efforts.

The focus of this chapter is on evaluation of broadening participation efforts, not on research on broadening participation, as important as that is. Since the goals of evaluation and research are different, this distinction is an important one. The major goal of research is to move the knowledge base forward, while for evaluation it is to assess the quality/effectiveness of a product, process, or project. In evaluation it can be enough to determine if something works and under what conditions; in research there is generally a why.

Research builds on the existing literature in the specific area studied, including theory. Even when exploratory research is done to generate hypotheses and/or theory, it must be built on something. In evaluation it is that which is being evaluated that should be based on existing research and theory. In research, the major focus is on the research, whereas in evaluation the major focus is on that which is being evaluated.

There are no differences between research and evaluation in terms of study design, measures, or analytic methods. The designs and procedures described in "Selecting and Implementing Appropriate Evaluation Designs," below, can be used for both evaluation of broadening participation programs and research on those efforts. There are, however, two methodological areas, longitudinal tracking and comparison groups, that have particular implications for the evaluation of broadening participation.

LONGITUDINAL TRACKING

> *Being able to follow students longitudinally is the key to any sophisticated understanding of how colleges are doing and what's happening to students.*
> Thomas R. Bailey, director of the Community College Research Center at Columbia University's Teachers College (Glenn, 2008, A10)

Bailey's comment about the importance of longitudinal tracking holds at the precollege and graduate levels as well as the college level. Indeed, longitudinal data may need to be extended to the workforce. Without such longitudinal data, the generation and testing of causal models tied to successful participation in STEM for diverse populations will be difficult if not impossible. While there has been interest in a national database, Congressional concern about privacy issues has made such a database difficult to enact at this time. A number of states are working on tracking students from when they enter school through college and perhaps beyond; however, most states are at early stages of development. There are a number of existing longitudinal data sets such as the National Center for Educational Statistics (NCES), the Early Childhood Longitudinal Study (ECLS), and the National Education Longitudinal Student (NELS); however, the variables for these studies have already been established and may not include what is needed to effectively study broadening participation (Clewell & Campbell, 2008).

The role of comparison groups in summative evaluations has been increasing in importance. The three types of study designs included in the U.S. Department of Education's *Report of the Academic Competitiveness Council* (2007, p. 14) Hierarchy of Study Design of Evaluating the Effectiveness of a STEM Education Intervention, by Expected Distribution of Study Type (Figure 6.1, below) all include some form of comparisons.

Figure 6.1: Three Types of Study Designs.

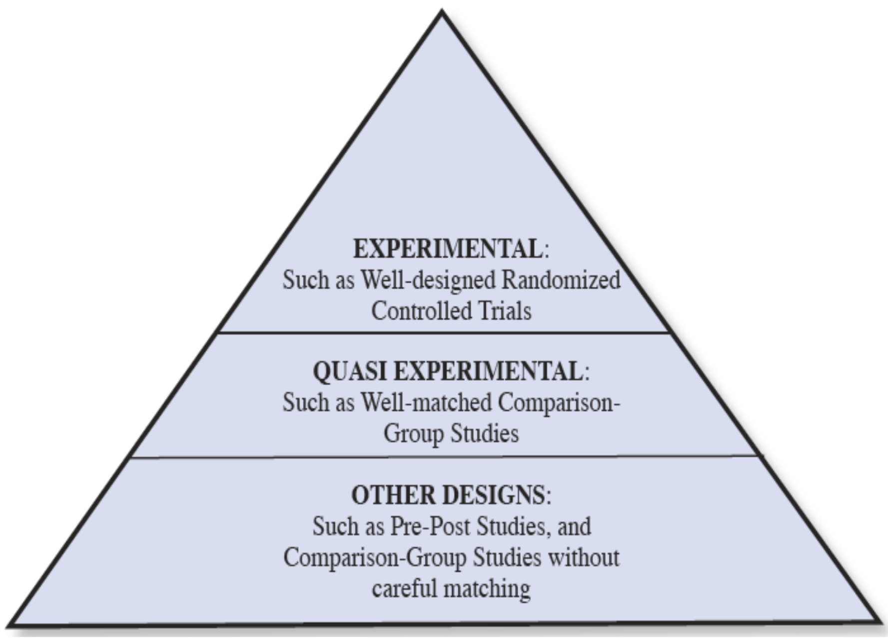


Figure 6.1. U.S. Department of Education's *Report of the Academic Competitiveness Council* (2007, p. 14) Hierarchy of Study Design of Evaluating the Effectiveness of a STEM Education Intervention, by Expected Distribution of Study Type

Obtaining adequate comparison groups for evaluation studies on broadening participation is not an easy task. There are public and political concerns about not offering services, even in the name of evaluation of impact, to groups in need. For example, a proposed experimental study of the Federal TRIO Programs, which are educational opportunity outreach programs designed to motivate and support students from disadvantaged backgrounds, was not allowed by Congress because it included a comparison group who would not be able to receive services.

Evaluators have looked to NSF collected data, including the Survey of Earned Doctorates (SED), as an important source of needed comparison data. Because of a concern about small cell sizes and confidentiality, in the past, the National Science Foundation’s (NSF) Division of Science Resources Statistics (SRS) has been suppressing data, which include the numbers of underrepresented minority Ph.D. recipients. The application of data suppression methods to NSF reports using the 2006 SED data compromised the utility of SED data for many data users. However, an alternative disclosure protection strategy incorporating the insights of several SED data user communities was developed in 2008-2009 to provide a better balance of confidentiality protection and data utility. The new disclosure protection procedures, which involve the aggregation of counts of doctorate recipients from small fields of degree, have been successfully applied to several reports and statistical tables that present 2007 and 2008 SED data. For a discussion of the origins of the SED reporting issue and the disclosure protection strategy implemented to resolve the issue, see http://www.nsf.gov/statistics/srvydoctorates/sedreporting/.

SELECTING AND IMPLEMENTING APPROPRIATE EVALUATION DESIGNS

Among the practical issues that those who design and commission evaluations of broadening participation programs commonly encounter are the following:

The Appropriateness of the Fit between the Design of the Program or "Intervention" and the Requirements of More Rigorous Evaluation Methodologies

The timing of the evaluation also has an impact on the design to be selected. Typically, more rigorous evaluation designs require certain compromises on the part of programs in order to facilitate key facets of the evaluation design. For instance, experimental design methodologies require the use of control groups from whom program services must be steadfastly withheld. Hence, one explicit tradeoff involves the rigor of the evaluation design versus denying the potential benefit of the program to some "worthy" individuals or groups. When these more rigorous methodologies are considered, a decision must be made about the relative value of pursuing higher levels of evidence of program effectiveness. Consideration must also be given to the suitability of a program (given its design) to facilitate such an evaluation. For example, consideration must be given to many factors such as whether program participants can be randomly assigned and whether there are enough participants to have a control and treatment group. Additionally, if groups have already been selected in a broadening participation project, then random assignment is not possible and if the intervention has already begun, pretest data will be limited to that which has already been collected.

The Balance between the Level of Investment in the Evaluation and the Level of Investment in and the Intensity of the Intervention (They Should be Roughly Commensurate)

Here, we are not speaking in terms of a balance between the financial investment in the intervention and the evaluation. Instead, this speaks to the balance between the intensity of the intervention and the rigor of the evaluation design. For example, it probably does not make sense to invest in an elaborate or lengthy evaluation to study the effects of a relatively light intervention such as one that exposes students to a limited set of information that is expected to help them (e.g., a one-day career fair designed to expose students from underrepresented groups to STEM careers). Conversely, it is quite worthwhile to consider the extent to which very intensive broadening participation interventions, applied on a large scale, warrant more rigorous evaluations to provide stronger information.

The Level of Evidence Expected Given the Nature of the Intervention

A STEM curricular or instructional intervention that makes fundamental changes to what is being taught or how instruction occurs, for instance, may be one that particularly warrants rigorous study employing methodologies suited to produce high levels of evidence. There may be many competing explanations for results and there may be opportunity costs associated with the forgone practices. In these instances, it is important to weigh the effectiveness of those traditional practices in relation to the new practices. Other types of interventions, like ones that promote change in institutional practices related to broadening participation (e.g., staffing, policies, programs and institutional commitment, accountability, and rewards), may not lend themselves well to rigorous measurements, and may produce changes that can be captured in less intensive studies.

The Strength of Rival Hypotheses

A rival hypothesis is a competing theory that might plausibly explain an outcome. As is noted above, a STEM curricular or instructional intervention enters a crowded environment, and findings might be subjected to alternate explanations for results. However, in cases where broadening participation programs involve offering substantial scholarship assistance or unique supplemental support, there may not be a strong rival hypothesis that could plausibly explain the outcome.

These are the types of considerations funding agencies and evaluation designers are weighing when selecting an appropriate evaluation approach for a program initiative. Evaluators of broadening participation programs, as is the case with other types of efforts, face many practical and methodological issues related to the selection and implementation of an appropriate design.

EVALUATION DESIGNS

What follows is a discussion of the major evaluation design options that may be appropriate for different types of broadening participation programs. The strengths and weaknesses of each type of evaluation design along with the feasibility of its successful implementation are provided. Attention is also devoted to identifying the factors that may enable or inhibit the successful application of the evaluation approach.

It is important to note that for any of the evaluation approaches, and in particular the more rigorous designs, a planning process that harmonizes program and evaluation designs from the outset, prior to the launch of an evaluation, can prove to be a very valuable facilitating factor. For the more rigorous evaluation methodologies to be used, as should become evident from the discussion below, serious consideration should be given to planning and initiating broadening participation programs and evaluations simultaneously. This allows for an integration of evaluation and program design. A common approach toward facilitating this is through the use of program and evaluation planning grants.

Experimental Designs

Optimally, a summative evaluation captures outcomes and links them to an intervention in a manner that establishes a causal relationship. When this is done well, it is possible to draw conclusions about a program's impact. This is done by separating out or "isolating" the effects of a particular intervention program. Experimental designs in evaluating broadening participation programs offer the most promise for establishing causal relationships.

The experimental design that is seemingly most commonly discussed is the randomized controlled trial (RCT). This approach can be summarized in the following manner:

R	O	X	O
R	O		O

Where *R* denotes random assignment; *O* denotes an assessment or measurement; and *X* denotes the intervention. Taken together, the symbols reflect that members of one group are randomly assigned to two groups; one of these groups participates in an intervention, while the other does not and both groups are subjected to pre-/post-assessments.

When this evaluation design fits (or can be made to fit) a broadening participation programmatic approach, it offers numerous methodological strengths which are valued by many evaluation stakeholders. Most notably, if implemented effectively, it rules out numerous rival hypotheses that could otherwise be seen as plausible causes of an outcome. It does so by studying the difference in the performance on an outcome measure between a group participating in a broadening participation intervention and the performance of a comparable group not exposed to the intervention. The random assignment procedure is generally seen as the best way to construct a truly comparable group for comparison purposes. Theoretically, if the groups were perfectly comparable, the only difference observed would be due to the intervention.

For the purposes of considering the application of this and subsequently discussed methodologies to broadening participation programs, we use the example of a prototype "pipeline program" that aims to increase the participation of women in engineering careers through the provision of supplemental support services (e.g., intensive academic guidance and mentoring) and through the provision of expanded and enhanced summer and school-year laboratory research experiences in baccalaureate programs. The ultimate goal of the prototype program would be increasing students' persistence in undergraduate degree programs and their degree attainment (possibly increasing their enrollment and degree attainment in graduate programs) and increasing their attainment of engineering positions and persistence in the field. As is discussed throughout this report these are reasonably typical goals for broadening participation programs.

If our prototypical intervention were implemented at one or more large polytechnic institutions, and such institutions were interested in participating in an experiment, it seems plausible to consider an experimental design evaluation approach. For instance, one or more freshmen classes of entering female engineering students could be randomly assigned to treatment and control groups. The treatment group would be exposed to the intervention, and the control group would experience the regular undergraduate program.

Feasibility of implementation. Many conditions would have to be in place to facilitate successful implementation of this type of experimental design effort. In the example provided above, the participating institution(s) of higher education would have to place value in the lessons learned from studying this broadening participation strategy. The funding agency would need to devote considerable resources and make a sustained commitment to an expensive evaluation approach and to the program strategy. The intervention would have to be thoroughly implemented and applied only to the treatment group; and the intervention could not spill over to the control group. (In practical terms, in our example for instance, this would mean engineering faculty who responded favorably to some of the supplemental services provided to the treatment group students would have to be careful not to offer those services to control group students.) Group sizes sufficiently robust to support analyses would be necessary for the evaluation to be carried out, and control and treatment group sizes would have to be sustained over the duration of a longitudinal project. In addition (as has been discussed throughout this document), a commitment on the part of all stakeholders would be needed to ensure the collection of quality longitudinal data.

It is easy to envision many of these conditions being hard to meet. Hence for this, the most difficult approach to undertake, it is certainly worth noting that it can be done. For example, evaluation of the Opening Doors project, conducted by MDRC, a nonprofit research group best known for mounting large-scale evaluations of real-world policies and programs targeted to low-income people, represents a good example of a rigorous evaluation with a randomized controlled trial. In Opening Doors, MDRC works with community colleges in several states to design and implement new types of financial aid, enhanced student services, and curricular and instructional innovations, with the goal of helping low-income students earn college credentials as the pathway to better jobs and further education. The evaluation uses a random assignment research design in which the experiences of students who receive the Opening Doors interventions are compared with those of students who receive existing services. The study tracks students for at least two years and measures the effects of Opening Doors on outcomes such as continued enrollment in college, academic performance, credential attainment, labor market success, and measures of individual well-being.

Quasi-Experimental Designs

These designs offer another set of methodological approaches for summative evaluations seeking to link program outcomes to broadening participation interventions. When these designs are executed effectively they provide information about program effects. The pretest/posttest version of quasi-experimental designs can be summarized as follows:

$$\begin{array}{ccc} O & X & O \\ O & & O \end{array}$$

Where *O* denotes the measurement or assessment; and *X* denotes the intervention. Taken together, the symbols reflect that members of a treatment and (non-randomly formed) comparison group are subjected to pre- and post-assessments on an outcome measure, but only the treatment group is exposed to the intervention.

Like the experimental design, quasi-experimental designs go to considerable lengths to isolate the effects of a program. This methodology, if executed well, can provide evidence of broadening participation program effects through comparisons of performance on an outcome measure between a group exposed to an intervention and a comparison group that has not been exposed to the intervention. In this way the logic that underlies this and the experimental design approaches are essentially the same. The strength of the quasi-experimental approach, however, in many ways depends upon the extent to which the comparison group is truly comparable to the group receiving the program services. It is generally difficult to construct good, matched comparison groups, and certain characteristics such as a self-selection bias may not be controlled for; randomization in theory takes care of such problems.

Nonetheless, when contrasted with the experimental design, an advantage of the quasi-experimental approach is that it is much less intrusive to the program. It is not necessary to assign students to a treatment or control group, and it is possible to employ quasi-experimental methodologies without denying students the new support and services offered through the broadening participation program.

This evaluation approach might be applied to our prototypical broadening participation program if the intervention were made available to an area of concentration (such as electrical engineering) within a participating polytechnic university and not in another (e.g., mechanical engineering) comparison group. Under such an approach, if the treatment and comparison groups (based upon matched characteristics or pretest measures) were deemed to be comparable, the difference between the two groups from pretest to posttest measurement would be expected to be related to the intervention.

Feasibility of implementation. A quasi-experimental design generally provides the most convincing alternative for assessing impact when randomized experiments cannot be conducted. The quasi-experimental methodology, which addresses rival hypotheses, makes a serious effort to construct comparison groups in an attempt to isolate the effects of the program and approximate a randomized design. In the example provided here, the program could operate without major alterations, although the possibility of the intervention spilling over to the comparison group would exist (as it did in the prior example), and would have to be guarded against. Thus, it is reasonably common for quasi-experimental designs to be employed where programs are already underway prior to the start of the evaluation. The complexity in such situations is in finding or constructing acceptable comparison groups by either matching participating and nonparticipating targets, or by performing statistical adjustments of participants and nonparticipants in an effort to make them equivalent on relevant variables. A valuable feature of the quasi-experimental methodologies is that they offer rigorous approaches toward studying the impact of programs that are underway, which would not offer the prospect of random assignment.

The quasi-experimental designs, like the experimental designs, are well-suited to situations where funding agencies and program operators and participants are seriously interested in studying (or demonstrating) the effectiveness of a strategy. The program and evaluation would have to be well-resourced, and a high level of sustained commitment on the part of all evaluation stakeholders would be required. Existing national databases may provide less costly alternative comparison groups. Clewell and Campbell (2008) have a list of such databases including descriptions of the sample, the variables, and how to access them.

Pre/Post Designs (Before/After Study)

Under this approach a summative evaluation tracks outcomes by comparing participants' performance on an outcome measure prior to a broadening participation intervention being implemented to their performance on that measure after implementation on the same targets. The post-measurement should be taken after sufficient time has elapsed for any effects to be expected. Some would argue that this is a quasi-experimental design because the premeasure is compared to the post, allowing the person to serve as their own comparison. This design can be summarized as follows:

O X O

Where *O* symbolizes an assessment; and *X* denotes the intervention. Taken together, the symbols reflect that members of a group are exposed to an intervention and subjected to pre- and post-assessments on an outcome measure.

This methodological approach does not address rival hypotheses. That is, it does not attempt to isolate the effects of the intervention, which is a limitation because it is always possible that something other than the intervention is causing any change detected on an outcome measure. Nonetheless, the approach does offer the promise of being able to capture whether change on a program outcome measure has occurred. Essentially this is accomplished by examining preprogram performance on an outcome measure and comparing it to the performance after a program has been implemented. If change on an outcome measure can be situated in a broader context (through the use of comparison samples), it can be easier to detect whether the change is meaningful.

As extension to the pre/post design is a pre/post design with comparisons or generic controls (e.g., established norms about typical changes in the target population). Applied to our prototypical engineering pipeline intervention, one could compare pre-project implementation data to post-implementation data on the number and percentage of women at a university persisting in and completing engineering programs. These data could be compared to program completion data from larger (possibly nationally representative) samples to situate the pre/post performance in a context.

A key consideration when deciding on the appropriateness of this approach pertains to the need to deal with rival hypotheses. Several of the authors of this document worked on a prior pipeline initiative evaluation project dealing with the recruitment of nontraditional teaching candidates for hard-to-staff urban and rural schools. In that effort, paraprofessionals working within the hard-to-staff schools were one of the populations targeted by the program intervention. Many of the paraprofessionals were long-term paraprofessionals and had associate's degrees and required an additional two years of schooling to meet teacher certification requirements. The intervention was multi-faceted, very intensive, and featured a lot of scholarship assistance. In that instance, it was deemed highly unlikely that the paraprofessionals would have become fully certified teachers in those hard-to-staff schools without the intervention. Hence investing in a complex methodology to provide a high level of evidence that the intervention as opposed to some competing explanation was responsible for moving the paraprofessionals through the pipeline seemed unwarranted, and a pre/post design was used.

Pre/post designs do not establish causality. However, as the example above suggests, in some situations other summative information about broadening participation program outcomes may suffice and prove quite useful. Pre/post designs do not interfere with the program operations (i.e., no one has to be denied treatment). It is also probably feasible to conduct such evaluations across a larger set of projects or an entire program initiative.

Feasibility of implementation. In most instances under a pre/post evaluation the program being studied could operate without alterations. A pre/post design can be employed most easily when a broadening participation program is starting. However, it can also be used for programs that are underway assuming that "pre-intervention" data are available or can be constructed. Also the comparisons that help contextualize a set of outcomes are clearly only feasible if data on appropriate comparison samples are accessible.

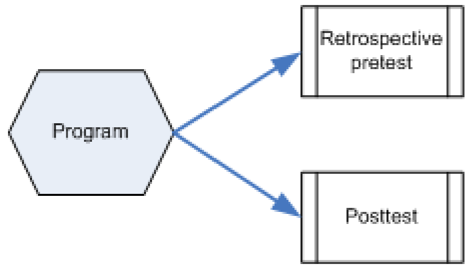
The pre/post (with comparisons) design will not be the first choice for evaluation stakeholders interested in definitively establishing whether or not a strategy is effective. That said, for some broadening participation programs, where there are not an array of plausible competing explanations, it may provide enough evidence of program effectiveness to be compelling to stakeholders and decision makers. In addition, it offers an option for studying whether desired outcomes are occurring and it offers an approach that can be applied on a larger scale than is practical for more complex evaluation approaches.

Retrospective Pretest Designs

The retrospective pretest (RPT) or "then-post" design can be a useful approach to collecting self-reported changes in knowledge, skills, intentions, behaviors, and attitudes of participants in broadening participation initiatives. This design is most useful when the results will be used in conjunction with other data. This relatively underutilized design is particularly useful when there are time or access constraints that allow the evaluator only a single opportunity to gather data from participants. A traditional pretest is administered before the program and again at the end of the program (posttest) with the notion that intervention effects are demonstrated by differences in the two measures. However, in some evaluations, use of the traditional pretest/posttest approach (with two data collection points) is not feasible and prohibitively expensive. In such situations, the RPT is a good alternative to the traditional pretest/posttest design and is a viable option to estimate pretest/posttest change (Lamb, 2005; Moore & Tananis, 2009).

In RPT designs, data are collected once at the end of the broadening participation effort. At that time, individuals are asked to assess their current level of knowledge/attitudes/skills/intentions *after* experiencing the intervention and to reflect on their previous level of knowledge/attitudes/skills/intentions *before* experiencing the broadening participation intervention. The RPT can be summarized as follows:

Figure 6.2: RPT Design Diagram.



A more robust RTP involves administering the retrospective pretest at a different time than posttest or at least in different parts of the questionnaire which, if done electronically, could prevent the editing of different responses (Taylor, Russ-Eft & Taylor, 2009). This approach reduces the response shift or other situational effects and limits bias related to effort justification (Hawkins, 2009).

On the positive side, the RPT design provides information about the intervention not available in posttest-only measures (Campbell & Stanley, 1963), and it is a fairly common practice in some areas of program evaluation (Hill & Betz, 2006; Moore & Tananis, 2009). Further, this design permits participants an opportunity to reflect upon how much they have changed as a function of their work in the intervention. For example, this design offers the opportunity to assess self-reported changes in students' attitudes, such as interest in STEM careers, and perceived usefulness of STEM in everyday life, after participating in broadening participation efforts. The RPT methodology also provides the evaluator with a more stable sample of participants who complete the instrument that can be used in data analysis by collecting responses for both measures at the same time in contrast to the traditional pretest/posttest design, where there is often a high drop-out of participants by posttest.

The RPT design is also viewed as a way to reduce the threat to validity due to response shift bias that often occurs in the traditional pre/post design (Howard, 1980; Moore & Tananis, 2009). This response shift results when individuals use a different frame of reference of understanding about a question between the pre- and post-periods. For example, individuals may not accurately assess their pre-intervention knowledge (often overestimating their initial level of competency) or skill; however, at the end of the intervention, their new understanding of the intervention content may affect their response on the post self-assessment.

Feasibility of implementation. The RPT clearly challenges traditional methodological logic, since both pretest data and posttest data are collected after the intervention has taken place. RPT designs are simple and cost-effective by reducing the costs and time required for data collection. This design is well-suited to STEM broadening participation evaluations interested in assessing changes on a range of cognitive, attitudinal, and skill-based variables in instances when the intervention has already begun before the evaluation begins, thereby eliminating the possibility of collecting baseline data. Further, RPT designs provide another data point that posttest-only designs do not provide.

A major drawback of the RPT design approach is that is does not address rival hypotheses. Similar with the (one group) traditional pretest/posttest design, RPT designs will not allow the evaluator to determine causality since this design does not include a control group of individuals who did not receive the intervention. Thus, the evaluator is unable to rule out threats to validity such as history, regression to the mean, and memory recall. Another limitation to the retrospective design is the recall period. For example, asking students to reflect back to the time prior to the broadening participation program can pose problems in terms of how accurately they can remember over time, with certain samples (e.g., elementary school-aged children) likely having more difficulty in accurately recalling pre-intervention knowledge, attitudes, behaviors, and skills than other samples.

Table 6.1: The Best Design for the Question.

Study Type	**Design**	**Representation**[1]	**Sample Broadening Participation Questions Answered by the Design**
Quantitative Case Study	One-Shot Posttest-Only Design	X O	After attending a preview weekend are at least 50% of the students planning to apply to the institution?
Quasi-Experimental Study	One-Shot Pretest/Posttest Design[2]	Oa X Ob	Does working with a role model increase girls' interest in science careers?
Quasi-Experimental Study	Retrospective (Post-Then-Pre) Design*	X Ob Oa	Does participating in a STEM bridge program increase students' perception of their current and past commitment to a STEM career?
Quasi-Experimental Study	Posttest-Only Intact Group Design	X O O	Do URM students participating in REU go on to graduate school in greater numbers than other URM students?
Experimental Study	Posttest-Only Design with Random Assignment	R X O R O	Similar questions as above, *R* increases comparability of comparison group.
Quasi-Experimental Study	Pretest/Posttest Intact Group Design	Oa X Ob Oa Ob	What is the impact of academic tutoring on student GPA?
Experimental Study	Pretest/Posttest Design with Random Assignment	R Oa X Ob R Oa Ob	Similar questions as above, *R* increases comparability of comparison group.

[1] X= intervention/treatment; O=assessment/effects; R=random assignment.

[2] There is disagreement among evaluators as to whether this design is a quasi-experimental study.

Table 6.1: The Best Design for the Question (continued).

Study Type	Design	Representation[3]	Sample Broadening Participation Questions Answered by the Design
Experimental Study	Solomon Four Group Design	R Oa X Ob R X Ob R Oa Ob R Ob	Similar questions as above, *R* increases comparability of comparison group; the groups without pretests account for any effect of testing.
Quasi-Experimental Study	Time Series Design	Oa Ob X Oc Od	Is time-to-degree for URM students declining?
Ethnography	Participant Observer Examination of Group Behaviors and Patterns	N/A	In what ways are the STEM academic experiences of female and male students different?
Case Study	Exploration of a Case (or Multiple Cases) Over Time	N/A	What institutional changes have occurred in the years that URM enrollment has been increasing?
Content Analysis	Systematic Identification of Properties of Large Amounts of Textual Information	N/A	Based on the paper trails, what are the similarities and differences in women and men's STEM tenure decisions
Mixed Methods Study	Use of More than One of the Above Designs	N/A	

Table 6.1: Table 6.1 provides a summary of the designs described above, as well as a number of other popular research designs. The table also provides specific examples of questions these designs might address in evaluations of broadening participation programs. Table adapted from: Campbell & Stanley, 1963; Ingersoll, 1983; Lydia's Tutorial Qualitative Research Methods, n.d.; Writing@CSU, ND.

[3] X= intervention/treatment; O=assessment/effects; R=random assignment

REFERENCES

Campbell, D.T. & Stanley, J.C. (1963). *Experimental and quasi-experimental designs for research on teaching.* Chicago: Rand McNally.

Campbell, P.B. & Clewell, B.C. (2008, January). *Building Evaluation Capacity: Guide II Collecting and Using Data in Cross-Project Evaluations*. Washington, DC: The Urban Institute. Retried from http://www.urban.org/UploadedPDF/411651_guide2.pdf

Carlson, L. (May 6, 2008). Personal communication.

Clewell, B. C. & Campbell, P. B. (2008, January). *Building Evaluation Capacity: Guide I Designing A Cross Project Evaluation*. Washington, DC: The Urban Institute. Retried from http://www.urban.org/UploadedPDF/411651_guide1.pdf

Colorado State University. *Writing@CSU*. Retrieved from http://writing.colostate.edu/index.cfm

Glenn, D. (2008, May 9). Huge databases offer a research gold mine—and privacy worries. *The Chronicle of Higher Education*, pp. A10. Retrieved from http://chronicle.com/article/Huge-Databases-Offer-a/24702

Hawkins, A. J. (2009). Retrospective deflation vs. situational inflation. *American Journal of Evaluation,* 30(3), 440-441.

Hill, L.G. & Betz, D.L. (2005). Revisiting the retrospective pretest. *American Journal of Evaluation,* 26(4), 501-517.

Howard, G.S. (1980). Response-shift a problem in evaluating interventions with pre/post self reports. *Evaluation Review*, 4(1), 93-106.

Ingersoll, G. (1982). Experimental methods. In Mitzel, H. (Ed.), *Encyclopedia of Educational Research (Fifth Edition)* (pp 624-631). New York, NY: The Free Press.

Lamb, T. (2005). The retrospective pretest: An imperfect but useful tool. *The Evaluation Exchange,* XI (2)

Lydia's Tutorial Qualitative Research Methods. Retrieved from http://www.socialresearchmethods.net/tutorial/Mensah/default.htm

Moore, D. & Tananis, C. A. (2009). Measuring change in a short-term educational program using a retrospective pretest design. *American Journal of Evaluation*, 30(2), 189-202.

National Science Foundation. *Download Revised 2006 Survey of Earned Doctorates Tables*. Retried from http://www.nsf.gov/statistics/srvydoctorates/2006/sed06data.cfm

Taylor, P. J., Russ-Eft, D. F., and Taylor, H. (2009). Gilding the outcomes by tarnishing the past: Inflationary biases in retrospective pretests. *American Journal of Evaluation,* 30(1), 31-43.

U.S. Department of Education. (2007). *Report of the Academic Competitiveness Council*. Washington, D.C.: Retrieved from http://www.ed.gov/about/inits/ed/competitiveness/acc-mathscience/index.html

APPENDIX A: THE AUTHORS

Fitzgerald B. Bramwell earned his B.A. degree in 1966 at Columbia University with Dr. Harry Gray, and his M.S. and Ph.D. degrees in physical chemistry at the University of Michigan with Dr. Julien Gendell and Dr. Thomas Dunn. Following graduate school in 1970, he joined ESSO Research and Engineering as a Research Scientist. The following year he accepted a faculty position at Brooklyn College of the City University of New York. He held a visiting professorship at the University of the West Indies, served as a Member of Technical Staff and as a consultant at AT&T Bell Laboratories, and as a consultant to the American Cyanamid Corporation. He returned to the faculty as a professor of chemistry after serving for six years as the University of Kentucky Vice President for Research and Graduate Studies. In that role he was the principal administrator for research and for the graduate school at the University of Kentucky. Prior to that appointment he was Dean of Graduate Studies and Research at Brooklyn College of the City University of New York, where he joined the faculty as an assistant professor of chemistry in 1971. He was promoted to professor of chemistry in 1980. From 1995 through June 2001 he served as a Trustee for the Southeastern Universities Research Association, as a member of the Executive Committee for the Council of Research Policy and Graduate Education of the National Association of State Universities and Land Grant Colleges, and on the Board of Directors of the Oak Ridge Associated Universities. He was honored for his career achievements as a research scientist in 2003 in *African Americans in Science and Invention* and in 1996 in *Distinguished African American Scientists of the 20th Century*. He has worked extensively with numerous historically black and minority serving colleges and universities to assist them in obtaining federal grants and contracts. Phillips Academy honored him in 2000 with the Claude M. Feuss Award for distinguished public service.

Patricia B. Campbell, Ph.D., President of Campbell-Kibler Associates, Inc, has been involved in educational research and evaluation with a focus on formal and informal science, technology, engineering, and mathematics (STEM) education and issues of race/ethnicity, gender, and disability since the mid-1970s. Her B.S., from LeMoyne College, is in Mathematics; her M.S., from Syracuse University, is in Instructional Technology; and her Ph.D., also from Syracuse University, is in Teacher Education. Dr. Campbell, formerly a professor of research, measurement, and statistics at Georgia State University, has authored more than 100 publications including co-authoring *Engagement, Capacity and Continuity: A Trilogy for Student Success*; *What Do We Know?: Seeking Effective Math and Science Education* and *Good Schools in Poor Neighborhoods: Defying Demographics, Achieving Success*. Dr. Campbell was a member of the U.S. Department of Education's Impact Review Panel and was part of the team involved in the development of the National Science Foundation publication *Infusing Equity in Systemic Reform: An Implementation Scheme*. She received the Betty Vetter Research Award from Women and Engineering Program Advocates Network (WEPAN) and the Willystine Goodsell Award from the American Educational Research Association.

Beatriz Chu Clewell is Senior Advisor and Founding Director, Program for Evaluation and Equity Research (PEER) at The Urban Institute in Washington, D.C., where she has worked since 1994. As an education policy researcher, her work has focused on factors that influence the educational attainment of underrepresented groups, especially in science, technology, engineering and mathematics (STEM) fields. Her undergraduate degree was in English Literature, with master's and Ph.D. degrees in Educational Policy, Planning, and Analysis. All three degrees are from Florida State University. From 1981 to 1994, Dr. Clewell was a senior research scientist at the Educational Testing Service. She also taught at the Universidad Simón Bolívar in Caracas, Venezuela, as well as middle school in Tegucigalpa, Honduras. From November 1999 to July 2000, on leave from the Urban Institute, she was Executive Director of the Commission on the Advancement of Women and Minorities in Science, Engineering and Technology Development (CAWMSET), a bipartisan Congressional commission based at the National Science Foundation (NSF). She has conducted over 60 evaluations, many of them large-scale, multi-site, multi-method in scope, including evaluations of many of NSF's broadening participation programs. Dr. Clewell was a member of the U.S. Department of Education's Impact Review Panel and currently serves on the National Academies' Congressionally-mandated Committee on Underrepresented Groups and the Expansion of the Science and Engineering Workforce Pipeline. She is also the author or co-author of several publications and a recipient of the Distinguished Scholar Award from the American Educational Research Association.

Dr. Darnella Davis, senior analyst at COSMOS Corporation, directs program evaluation activities in the social sciences. Dr. Davis (Muscogee Creek) has conducted comprehensive and cross-site evaluations of education and youth development programs sponsored by public and private agencies and organizations. Often developing in-depth case studies, Dr. Davis has identified promising practices and made recommendations to practitioners, administrators, and policymakers. These studies have enabled Dr. Davis to pursue her interest in understanding the role of collaborative support systems in improving conditions for underserved populations. Dr. Davis recently completed an assessment of the National Science Foundation's investments in capacity building in evaluation through broadening participation.

Dr. Norman L. Fortenberry is the founding Director of the Center for the Advancement of Scholarship on Engineering Education (CASEE) at the National Academy of Engineering (NAE). CASEE facilitates research on and deployment of innovative policies, practices, and tools designed to enhance the effectiveness and efficiency of systems for the formal, informal, and lifelong education of engineers. He previously served in various executive positions within the National Science Foundation's Directorate for Education and Human Resources. He has also served as executive director of the National Consortium for Graduate Degrees for Minorities in Engineering and Science, Inc. (The GEM Consortium) and as a faculty member in the department of mechanical engineering at the Florida A&M University–Florida State University College of Engineering. Dr. Fortenberry was awarded the S.B., S.M., and Sc.D. degrees (all in mechanical engineering) by the Massachusetts Institute of Technology.

Dr. Antonio García is a Professor of Bioengineering in the Fulton School of Engineering at Arizona State University. He received a B.S. in Chemical Engineering, Rutgers University–New Brunswick in 1981 and a Ph.D. in Chemical Engineering, University of California, Berkeley, 1988. He is the Associate Director of the Hispanic Research Center.

Dr. Donna Nelson, associate professor of chemistry at Oklahoma University (OU), obtained her Ph.D. in chemistry at the University of Texas at Austin with M. J. S. Dewar, did her postdoctorate at Purdue with H. C. Brown, and joined OU in 1983. She researches three global challenges—energy, environment, and scientific workforce development—and frequently speaks on their interrelationship. She has over 90 publications and several honors, including Fulbright Scholar, NSF ADVANCE Leadership Award, SACNAS Distinguished Scientist, Women's eNews' 21 Leaders for the 21st Century, AAAS Fellow, Guggenheim Award, NOW Woman of Courage, Ford Fellow, Sigma Xi Faculty Research Award, NSF Creativity Extension, and many keynote talks. Her chemical research involves functionalizing single walled carbon nanotubes (SWNTs), with applications in energy research and technology development, and yielded the first COSY NMR spectrum of covalently functionalized SWNTs. Her scientific workforce surveys, of faculty race/ethnicity, gender, and rank in science and engineering at research universities, revealed that women and minorities are much less represented among professors than degree recipients. The Nelson Diversity Surveys final report is at http://cheminfo.chem.ou.edu/~djn/diversity/briefings/Diversity%20Report%20Final.pdf. More information is at http://cheminfo.chem.ou.edu/faculty/djn/djn.html.

Adam Stoll has been with the Congressional Research Service (CRS) since November 1999. Currently he is the Section Head of the Education and Labor Section in the Domestic Social Policy Division at CRS. He leads a group of 12 analysts in the development of research and policy analysis to support the needs of the Congress on issues pertaining to education, workforce development, and the workforce. During most of his tenure at CRS, Adam has served as lead analyst on all issues pertaining to the federal student loan programs. He also worked as a senior analyst on projects examining the federal need analysis system (which is used to determine student eligibility for various sources of financial aid), and on projects examining the relationship between higher education tax credits and traditional student aid. Immediately prior to joining CRS, Adam worked as an evaluation officer at the Lila & DeWitt Wallace Reader's Digest Funds. In that position, he designed evaluations for large national demonstration projects dealing with an array of issues, including: teacher recruitment, teacher professional development, comprehensive school reform, adult literacy, and urban park creation and revitalization. In earlier career stops Adam worked for the American Educational Research Association and for the Center for Education and the American Economy at Teachers College–Columbia University. Adam has a Ph.D. in Social Policy from the Heller School, Brandeis University.

Veronica G. Thomas is a Professor in the Department of Human Development and Psychoeducational Studies at Howard University. She is also a Senior Research Associate with the Capstone Institute, Howard University. Dr. Thomas was the principal investigator of the Howard University Evaluation Training Institute, a National Science Foundation (NSF) funded project, and the former co-principal investigator of the Secondary School Project at Howard University's Center for Research on the Education of Students Placed at Risk (CRESPAR). Her research interests include the academic and socio-emotional development of youths placed at risk, well-being of Black women and girls, and culturally responsive evaluations. Dr. Thomas has authored or co-authored work in venues such as *New Directions for Evaluation*, *Adolescence, Educational Leadership*, *Journal of Adult Development*, *Review of Research in Education*, *Journal of Negro Education*, *Family Relations*, *Journal of Black Psychology*, *Sex Roles*, *Journal of Social Psychology*, *Women and Health*, and the *Journal of the National Medical Association*. Her major professional associations include the American Psychological Association (APA), the American Evaluation Association (AEA), American Educational Research Association (AERA), and the Eastern Evaluation Research Society (EERS). Dr. Thomas has served as an evaluation consultant or trainer for various projects/organizations, including the National Council on Community and Education Partnerships (NCCEP), Northern Virginia Resource Mothers' Program, Council of Graduate Schools' Preparing Future Faculty (PFF) Program, District of Columbia Public Schools' HIV/AIDS Education Program, Urban Family Institute's Kids House Program, Minority Graduate Education Program, Center for Substance Abuse Treatment (CSAT), and the Center for Substance Abuse Prevention (CSAP).

APPENDIX B: WORKSHOP AGENDA

Workshop on Evaluation of Efforts to Broaden Participation in STEM

April 17-18, 2008 ~ Arlington Hilton Hotel ~ Gallery II ~

950 North Stafford Street, Arlington, VA 22203

Thursday, April 17, 2008

		Location
7:30 – 8:00 a.m.	Registration & Light Refreshments	Gallery II
8:00–10:00 a.m.	Welcome and Opening Keynote **Bernice Anderson**, *Senior Advisor, EHR/OAD* **Beatriz Clewell**, *Director, Program for Evaluation and Equity Research (PEER), The Urban Institute* **Norman Fortenberry**, *Director, Center for the Advancement of Scholarship on Engineering Education, National Academy of Engineering* "NSF Perspective on the Importance of Broadening Participation" **Margaret E.M. Tolbert**, *Senior Advisor, OIA and CEOSE Executive Liaison National Science Foundation*	
10:00 – 10:30 a.m.	Break	

Workshop on Evaluation of Efforts to Broaden Participation in STEM

April 17-18, 2008 ~ Arlington Hilton Hotel ~ Gallery II ~

950 North Stafford Street, Arlington, VA 22203

Thursday, April 17, 2008 (continued)

		Location
10:30 – 12:00 a.m.	Parallel Sessions Break outs meet to discuss the topics (a) metrics for project monitoring and (b) program evaluation design and indicators	Gallery I,II, & III
12:00 – 1:00 p.m.	Report Outs and Discussion	Gallery II
1:00 – 2:30 p.m.	Working Lunch Guest Speaker, **Joel Parriot**, *Office of Management and Budget:* "A View from the Office of Management and Budget"	Gallery II
2:30 – 4:30 p.m.	Parallel Sessions Draft written recommendations on metrics and evaluation design	Gallery I, II, & III
4:30 – 5:30 p.m.	Report Outs **Beatriz Clewell** and **Norman Fortenberry**	Gallery II
5:30 p.m.	Adjourn for Day **Beatriz Clewell** and **Norman Fortenberry**	

Workshop on Evaluation of Efforts to Broaden Participation in STEM

April 17-18, 2008 ~ Arlington Hilton Hotel ~ Gallery II ~

950 North Stafford Street, Arlington, VA 22203

Friday, April 18, 2008

		Location
8:00 – 8:30 a.m.	Light Refreshments	Gallery I
8:30 – 8:45 a.m.	Workshop Reconvenes	Gallery I
8:45 – 9:15 a.m.	Synthesis of Thursday's Results **Beatriz Clewell** and **Norman Fortenberry**	Gallery I
9:15 - 9:30 a.m.	NSF Perspective on Importance of Workshop: **Celeste Rolfing**, *MPS/CHE, Co-Chair, NSF Working Group on Broadening Participation*	Gallery I
9:30 -10:00 a.m.	Break	Gallery I
10:00 a.m. - 12:30 p.m.	Parallel Sessions:	Gallery I
	Groups look at implementation issues – brainstorm on possible objections/concerns and how to address same. Draft recommendations.	DaVinci Picasso Matisse
12:30p.m.	Closing Remarks, Workshop Adjourns: **Elmima Johnson**	Gallery I

Workshop on Evaluation of Efforts to Broaden Participation in STEM

April 17-18, 2008 ~ Arlington Hilton Hotel ~ Gallery II ~

950 North Stafford Street, Arlington, VA 22203

Friday, April 18, 2008 (continued)

		Location
1:30 – 5:30 p.m.	Expert Panel Members Reconvene: **Beatriz Clewell** and **Norman Fortenberry** Work on outlining and writing final document	Gallery I
5:30 p.m.	Expert Panel Adjourns	

APPENDIX C: FOLLOW-UP MEETING AGENDA

Report on Workshop on Evaluation of Efforts to Broaden Participation in STEM

December 17, 2008 ~ Arlington Hilton Hotel ~ Gallery II ~

950 North Stafford Street, Arlington, VA 22203

Wednesday, December 17, 2008		
		Location
8:30 – 9:00 a.m.	Registration & Light Refreshments	Gallery II
9:00–9:15 a.m.	Welcome and Opening Keynote **Cora Marrett**, *Senior Advisor, Assistant Director, Education and Human Resources (EHR), NSF*	
9:15 – 9:30 a.m.	History of Effort **Bernice Anderson**, *Senior Advisor, EHR/OAD*	
9:30 – 11:00 a.m.	Chapter Summaries **Patricia B. Campbell**, *President of Campbell-Kibler Associates, Inc.* **Norman L. Fortenberry**, *Director, Center for the Advancement of Scholarship on Engineering Education, National Academy of Engineering* **Veronica G. Thomas**, *Professor, Department of Human Development and Psychoeducational Studies, Howard University* **Adam Stoll**, *Congressional Research Service*	Gallery II

Report on Workshop on Evaluation of Efforts to Broaden Participation in STEM

December 17, 2008 ~ Arlington Hilton Hotel ~ Gallery II ~

950 North Stafford Street, Arlington, VA 22203

Wednesday, December 17, 2008 (continued)		
		Location
9:30 – 11:00 a.m.	Chapter Summaries (continued) **Elmima Johnson**, *Program Director, REESE, Division of Research on Learning, EHR*	Gallery II
11:00 – 11:45 a.m.	Response to Report **Fae Korsmo**, *Senior Advisor, Office of the Director, NSF* **Beverly Karplus Hartline**, *Associate Provost for Research and Dean of Graduate Studies, University of the District of Columbia; Former Chair, CEOSE* **Carl Person**, *Manager, Minority University Research and Education Programs, Office of Education, NASA Headquarters*	Gallery II
11:45 – 12:00 p.m.	Audience Feedback **Norman Fortenberry** *Audience is invited to question authors and provide feedback*	Gallery II
12:00 – 12:15 p.m.	Next Steps to Include Dissemination **Bernice Anderson**	Gallery II
12:15 p. m.	Adjourn for Day	