# Building Evaluation Capacity

*Designing a Cross-Project Evaluation*

GUIDE I

Beatriz Chu Clewell

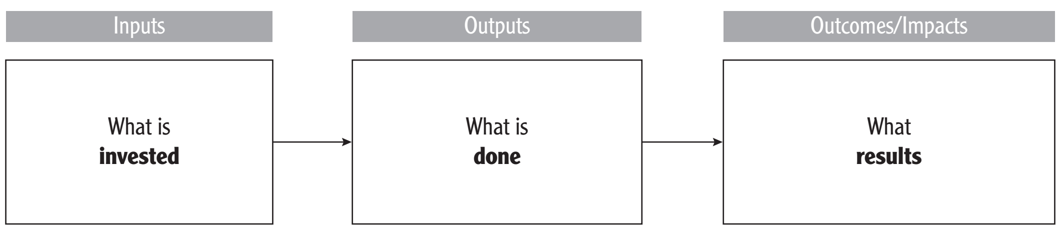
Patricia B. Campbell

JANUARY 2008

# Building Evaluation Capacity

*Designing a Cross-Project Evaluation*

GUIDE I




Beatriz Chu Clewell and Patricia B. Campbell

JANUARY 2008

THE URBAN INSTITUTE

The Urban Institute's Program for Evaluation and Equity Research (PEER) focuses on education research. PEER staff conduct studies in the fields of educational attainment, educational access, minorities in mathematics and science, teacher education, teacher recruitment and retention, and educational assessments.

This report was prepared for the National Science Foundation Directorate for Education and Human Resources (EHR) and supported by REC-0231644.

Copyright © February 2008. The Urban Institute. All rights reserved. Except for short quotes, no part of this report may be reproduced or utilized in any form or by any means, electronic or mechanical, including photocopying, recording, or by information storage or retrieval system, without written permission from the Urban Institute.

The Urban Institute is a nonprofit, nonpartisan policy research and educational organization that examines the social, economic, and governance problems facing the nation. The views expressed are those of the authors and should not be attributed to the National Science Foundation or to the Urban Institute, its trustees, or its funders.

# CONTENTS

## Acknowledgments

The authors wish to acknowledge the assistance of the many evaluator colleagues, agency staff, and grantees who contributed to these guides. The nine members of the Delphi panel, expert evaluators all, who participated in the process of identifying key indicators and measures for the selected programs, deserve our heartfelt thanks. Because one of the characteristics of the Delphi technique is that panel members retain their anonymity, we cannot reveal their identities. Program officers in the Division of Human Research Development at the National Science Foundation, the Division of Minority Opportunities in Research at the National Institutes of Health, and the GE Foundation provided valuable input and feedback to the guides. Principal investigators and project staff of the grantee projects featured in the guides—ADVANCE, AGEP, LSAMP, MARC, MBRS, and Math Excellence—participated in interviews and focus groups that gave us an on the ground look at the cross-project evaluation process from the perspective of the grantees. Marsha Matyas, of the American Physiological Society, and Jason Kim, of Systemic Research, advisors to the project, guided our early efforts. Bernice Anderson and Elmima Johnson, of the National Science Foundation, reviewed the draft guides and were instrumental in helping us enhance their usefulness to funding agencies. Kathy Zantal-Weiner reviewed the final draft of the manuscript and provided good suggestions for improving the guides' appeal to evaluators.

Last, but not least, we wish to thank the members of our project team: Matthew Buettgens, Laurie Forcier, Nicole Deterding, and Nicole Young, all formerly of the Urban Institute and Lesley Perlman and Rosa Carson, formerly of Campbell-Kibler Associates, who contributed substantively to the development and production of the guides. Will Bradbury of the Urban Institute was, as always, extremely helpful in overseeing the editing and production process.

# Building Evaluation Capacity

## Designing a Cross-Project Evaluation—Guide I

The national trend toward increasing accountability of the public sector, which gave rise to the enactment in 1993 of the Government Performance and Results Act (GPRA), has placed added emphasis on the ability of public agencies to report accurately and consistently on performance and results. The more recent No Child Left Behind Act, with its calls for scientific evidence in education research, and the current Campbell Collaborative, which seeks to improve social science research through a focus on random assignment in studies in social, behavioral, and educational arenas, indicate this trend is growing. Evaluation capacity building (ECB), as a system for enabling organizations and agencies—both public and private—to develop the mechanisms and structure to facilitate evaluation to meet accountability requirements, is a concept whose time has come.

One important way ECB differs from mainstream evaluation is that ECB is continuous and sustained rather than episodic. Other characteristics are that it is context-dependent; operates on multiple levels; is flexible in responding to multiple purposes, requiring continuous adjustments and refinements; and requires a variety of evaluation approaches and methodologies (Stockdill, Baizerman, and Compton 2002).

The goal of this project was to develop a model to build evaluation capacity in three organizations: the National Science Foundation (NSF), the National Institutes of Health (NIH), and the GE Foundation. More specifically, the project's intent was to test the feasibility of developing models to facilitate the collection of cross-project evaluation data for programs within these organizations that focus on increasing the diversity of the STEM workforce. To facilitate the standardization of the model, we chose programs with a similar goal: increasing the participation and success of underrepresented groups[1] in STEM fields of study and careers. A description of each program used to test the model appears in appendix A.

> ECB [Evaluation Capacity Building] is the intentional work to continuously create and sustain overall organizational processes that make quality evaluation and its uses routine. (Stockdill, Baizerman, and Compton 2002, 14).

## Programs within Three Funding Agencies

**National Science Foundation:**

*ADVANCE*
Division of Human Resource Development (HRD)

*Alliances for Graduate Education in the Professoriate (AGEP)*
Division of Human Resource Development (HRD)

*Louis Stokes Alliances for Minority Participation (LSAMP)*
Division of Human Resource Development (HRD)

**National Institutes of Health:**

*Minority Access to Research Careers (MARC)*
Division of Minority Opportunities in Research (MORE)

*Minority Biomedical Research Support (MBRS)*
Division of Minority Opportunities in Research (MORE)

**GE Foundation:**

*Math Excellence*

# Why Cross-Project Evaluation?

Several government agencies as well as private foundations target broad programmatic areas for funding. For example, NSF's Division of Human Resource Development (HRD) within the Education and Human Resources (EHR) Directorate, strives to address NSF's agency-wide commitment to broaden the participation of underrepresented groups and institutions in STEM fields. Within NIH, the National Institute of General Medical Science's (NIGMS) Division of Minority Opportunities in Research (MORE) supports research and research training programs whose goal is to increase the number of minority biomedical and behavioral scientists. Within these two broad programmatic areas—HRD and MORE—several programs are funded that address the same or similar goals and the same or similar target populations (i.e., AGEP, LSAMP, ADVANCE at NSF, MARC and MBRS at NIH).

In this era of accountability, programs are frequently asked to provide evidence that they are reaching their goals. Assessing broad program effectiveness, however, is often difficult without being able to aggregate similar data across individual projects to determine the overall effect of the program. Cross-project evaluation in the interest of evaluating a program depends on each project collecting uniform data so that data may be aggregated across projects.

In addition to meeting accountability requirements, being able to report on program level outcomes—*Has this program produced the intended results?*—brings other benefits. For example, it strengthens the ability of funders to make informed funding decisions—*Should we continue to put money into this program?*—and it has the potential to improve the knowledge base of what works (and what works for whom) by providing a larger database allowing for disaggregation of data—*Who benefits the most from this intervention? Who doesn't?*

The goal of cross-project evaluation is most often to provide data for a summative evaluation by assessing the program's overall success in meeting its goals as measured by the success of individual projects in contributing to those goals. This guide, therefore, will focus on the role of cross-project evaluation in summative evaluation.

The two guides target three groups of potential users:

- *Funders of program evaluations.* These can be either staff in the program area or staff in an evaluation center within an organization. They are responsible for writing requests for proposals (RFPs) to solicit proposals to evaluate programs, monitoring the evaluations, and working with the evaluators to convey the results of the evaluation to the organization and the public.
- *Those involved in program evaluation.* These are typically evaluators who receive contracts from funding agencies or organizations to design and carry out cross-project evaluations within a program to assess the overall effectiveness of the program.
- *Those involved in project evaluation.* These project evaluators focus on designing and collecting data to evaluate individual projects. They work directly with the projects and typically report their data to the project director or principal investigator.

These guides show how to develop a model to guide procedures for building the capacity and ability of an organization (or an area within an organization) to conduct cross-project evaluations. Evaluators responsible for assessing the effects of a program area can also use the guides to conceptualize cross-project evaluation and organize data collection and analysis. Finally, they can help evaluators of individual projects understand their role in contributing to the larger, cross-project evaluation.

There are two basic assumptions that underlie the evaluation approach used and the evaluation knowledge and expertise the targeted audience possesses. The evaluation approach (i.e., the goals and procedures of evaluation) presented here is the goal-oriented approach (Stecher and Davis 1990), which determines program success based on fulfillment of program-specific goals. We deemed this to be the most appropriate evaluation approach because of the purposes for which cross-project evaluations are typically used: to justify or inform funding decisions. For example, a program of funding is established to address an identified need through implementation of a set of projects that, by carrying out specified activities, promise to produce specific results. Cross-project evaluation, by determining whether the projects have achieved these results, justifies—or not—the decision to establish a funding initiative. The question the evaluation addresses—*Has this program fulfilled its goal?*—is therefore an appropriate one.

Because the target audience consists of evaluators, the guides assume that readers will have a basic knowledge of evaluation principles and procedures. The guides, therefore, do not contain detailed instructions for implementing routine evaluation tasks.[2] The expectation is that users of these guides will gain an appreciation of the value of conducting cross-project evaluation; learn how to design such an evaluation; become aware of the special issues inherent in collecting and reporting cross-project data; and understand the many uses of cross-project data in decisionmaking.

The two guides will discuss the following areas relevant to cross-project evaluation:

Building Evaluation Capacity—Guide I: Designing a Cross-Project Evaluation

- Evaluation design, including identification and operationalization of program goals, building of logic models, and indicator setting; and
- Selection of indicators to be measured and appropriate measures for these indicators.

Building Evaluation Capacity—Guide II: Collecting and Using Cross-Project Evaluation Data

- Data collection: formats and scheduling,
- Ensuring data quality,
- Use of comparison data,
- Reporting and displaying data, and
- Using cross-project data.

Designing a Cross-Project Evaluation is the first of the two building evaluation capacity guides. It focuses on evaluation design, including identification and operationalization of program goals, building of logic models, and indicator setting. This guide also covers the selection of indicators to be measured and appropriate measures for these indicators. It begins with the identification of program goals. Subsequent sections discuss the construction of logic models and the evaluation approach, including the generation of evaluation questions, the setting of indicators, and the integration of evaluation questions and indicators. It concludes with a discussion of measurement strategies: the selection of appropriate measures for different indicators and the role of demographic variables.

## Identifying Program Goals

The first step in designing a cross-project evaluation is determining the programmatic goals. For example, programs sponsored by NSF's Division of Human Resource Development (HRD), which is located in the Directorate for Education and Human Resources, focus on carrying out HRD's broad mandate, as expressed in the following statement from the NSF web site:

> The Division of Human Resource Development [HRD] . . . serves as a focal point for NSF's agency-wide commitment to enhance the quality and excellence of science, technology, engineering, and mathematics (STEM) education and research by broadening the participation and advancement of minority-serving institutions, women and girls, and persons with disabilities at every level of science and engineering enterprises including underrepresented minorities.

Within HRD, broad programmatic areas have been designated that focus on specific underrepresented groups and institutions that serve these groups: minorities and minority-serving institutions, women and girls, and persons with disabilities. Within the minorities and minority-serving institutions area are six programs: Historically Black Colleges and Universities Undergraduate Program (HBCU-UP), Tribal Colleges and Universities Program (TCUP), Centers for Research Excellence (CREST), the Louis Stokes Alliances for Minority Participation (LSAMP), Alliances for Graduate Education in the Professoriate (AGEP), and Model Institutions for Excellence (MIE). Each program addresses the HRD goals in a different way; some focus on a specific underrepresented population, such as TCUP (American Indians) or a specific academic level, such as LSAMP (undergraduate) and AGEP (graduate) for all underrepresented minorities (URMs). The two HRD programs in this category that our project selected are LSAMP and AGEP. See the text boxes below for the goals of these programs.

## Goal-Setting as a Process

Once the broad goals of the program are identified, what steps are required to

- refine broad goals?
- obtain consensus from stakeholders?
- operationalize refined goals?

*Refining Broader Goals.* In preparation for working with stakeholders to come to consensus on the definition and operationalization of programmatic goals, it is first necessary to refocus and refine the broader goals as conceptualized by the program. Again, following through with our example from the two HRD programs, the text boxes below show how broad LSAMP and AGEP program goals were refined for presentation to stakeholders.

This refinement of broad programmatic goal statements into “bite-sized” segments facilitates discussion with stakeholders in the next steps in the goal-setting process.[3] It is important to remember in the refining process that the integrity of the program must be maintained.

*Arriving at Consensus with Stakeholders.* Why should stakeholders be involved in goal-setting? First, stakeholders can help to illuminate the links between goals and program activities. They can also provide insight into and expand upon the stated goals of a project. (“Even though the goal statement only mentions impact on students, the program aims to improve curriculum and faculty instruction.”) Stakeholders can also provide a realistic perspective for attaining goals. (“There is no way this program can double the number of graduates, but it can increase them by 50 percent.”) Stakeholder consensus on what the program goals mean is important because it lays the groundwork for acceptance of the evaluation findings by the people who count. Because the goals are the departure point for the evaluation design, evaluators must ensure that goals reflect the understanding and acceptance of important stakeholders.

The evaluator must decide which stakeholders to involve. Because the process can be labor intensive and lengthy, this decision is critical.[4] In determining the stakeholders to be involved in this project, we focused on the primary users of the guide—the

### LSAMP

***Broad Goal***

The LSAMP Program is designed to develop the comprehensive strategies necessary to strengthen the preparation and increase the number of minority students who successfully complete baccalaureates in STEM fields. This objective facilitates the long-term goal of increasing the production of doctorates in STEM fields, with an emphasis on entry into faculty positions.

***Refined Goals***

- Increase the quality and quantity of URM students successfully completing STEM baccalaureate degree programs.
- Increase the number of URM students interested in, academically qualified for, and matriculated into programs of graduate study in STEM.

### AGEP

***Broad Goal***

The AGEP Program seeks to significantly increase the number of URMs receiving doctoral degrees in STEM fields customarily supported by NSF. NSF is particularly interested in increasing the number of minorities who will enter the professoriate in these disciplines.

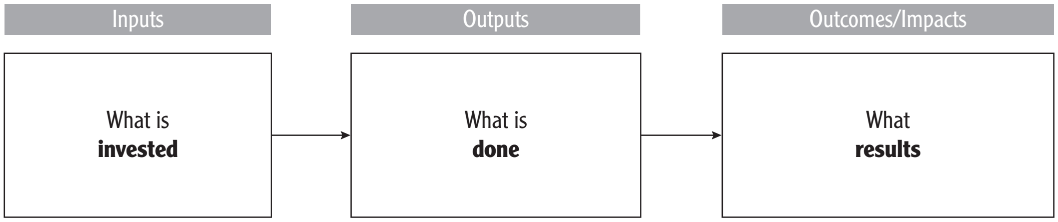
***Refined Goals***

- Increase the number of underrepresented minority students pursuing advanced study, obtaining doctoral degrees, and entering the professoriate in STEM disciplines.
- Establish alliances engaged in comprehensive institutional cultural changes that will lead to sustained increases in the conferral of STEM doctoral degrees, significantly exceeding historic levels of performance.

agency or foundation evaluation and program staff and experienced evaluators who were frequently contracted by the agencies and foundations to conduct program evaluation. We included a third group of stakeholders: project directors and principal investigators of the individual projects and their evaluators. These stakeholders, who conduct the data collection efforts in the field, must understand and buy into cross-project evaluation for it to work. Additionally, this group can provide valuable judgment regarding the feasibility of collecting certain types of data that might be proposed in an evaluation plan.

*Operationalizing the Goals.* This is the process whereby goals are transformed into expected outcomes, evaluation questions, and, ultimately, into indicators that can be measured quantitatively and qualitatively. Stakeholder input was used extensively in this process via methods described below. Lists of indicators for each outcome were generated and these were discussed with stakeholders, who identified the most important ones. How is the transition made from goals to measures of their attainment? A useful tool for aligning goals with indicators with measures is the logic model.

**Figure 1.** Components of a Logic Model




# Constructing a Logic Model

A logic model is the simplified, idealized, graphic depiction of a program or project.[5] A good logic model shows the logical relationships among resources invested, activities implemented, and the changes that result; in other words, a logic model illustrates how a program is supposed to work. Although thought of mainly as an evaluation tool, the logic model has been used in program planning and management. This section, however, explores the role of the logic model as a tool for evaluation.

## Components of a Logic Model

As one of the first steps in designing an evaluation, the logic model helps to focus an evaluation by identifying what needs to be evaluated. The basic components of a simplified model, as figure 1 depicts, are inputs (resources invested), outputs (activities implemented using the resources), and outcomes/impact (results). Figure 1 illustrates how these components can develop into a complex picture of a specific program.

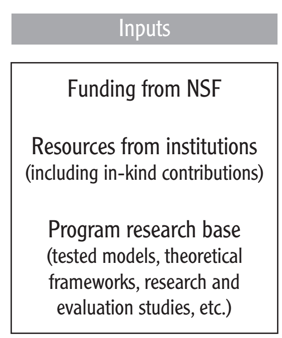
Let's apply our understanding of the logic model components to a specific program, LSAMP, and consider what goes into a logic model for this program, component by component.

*Inputs.* The inputs for LSAMP consist of funding from NSF, resources (including in-kind contributions) from the participating institutions, and a research base on which the program approach is based. Figure 2 gives inputs for the LSAMP Program.

*Outputs.* These consist of two types: activities (what the program does) and participants (who the program reaches). Figure 3 shows the outputs identified for the LSAMP Program.

*Outcomes/Impact.* There are three types of outcomes: short-term, medium-term, and long-term. Short-term outcomes occur within a few years after program initiation; medium-term outcomes may be evident about 10 years into the program; and long-term outcomes may be expected only after several years (figure 4).

**Figure 2.** LSAMP Model Inputs




## The Logic Model in Evaluation

The logic model has been widely used in evaluation for the following reasons: it helps to (a) determine the focus of the

**Figure 3.** LSAMP Model Inputs and Outputs

| Inputs | Outputs | |
|---|---|---|
| | **Activities** | **Participants** |
| Funding from NSF | Establish partnerships | Partner institutions |
| Resources from institutions (including in-kind contributions) | Conduct student professional-ization activities | Student participants |
| Program research base (tested models, theoretical frameworks, research and evaluation studies, etc.) | Improve curriculum | Faculty participants |
| | Develop faculty | |

**Figure 4.** LSAMP Model Inputs, Outputs, and Outcomes

| Inputs | Outputs | | Outcomes/Impacts | | |
|---|---|---|---|---|---|
| | **Activities** | **Participants** | **Short-term** | **Medium-term** | **Long-term** |
| Funding from NSF | Establish partnerships | Partner institutions | Student achievement in STEM | Student enrollment into a STEM graduate program | Student entry into STEM careers |
| Resources from institutions (including in-kind contributions) | Conduct student professional-ization activities | Student participants | Student graduation from STEM majors | Student attainment of a graduate degree in STEM | Institutionalization of LSAMP practices at participating institutions |
| Program research base (tested models, theoretical frameworks, research and evaluation studies, etc.) | Improve curriculum | Faculty participants | Curricular reform | Revisions to institutional policies and practices | Contributions to scholarly body of work on effective models for increasing minority participation in STEM |
| | Develop faculty | | Improvement of STEM faculty teaching | | |

## Uses of the Logic Model in Evaluation

1. Focus—Determine what to evaluate.
2. Questions—Determine the questions your evaluation will ask.
3. Indicators—Decide what data to collect to answer your evaluation questions.
4. Timing—Decide when to collect which data.
5. Data Collection—Determine data sources, methods, instrumentation, and samples.

evaluation—what will be evaluated; (b) determine the main evaluation questions—what should be measured; (c) understand indicators and determine what data best answer the evaluation questions; and (d) identify the best schedule for data collection.

## Evaluation Focus and Questions

As discussed in a previous section of this report, an assumption of the guide is that the evaluation approach to be used is the goal-oriented approach. The main evaluation questions, therefore, should be developed to correspond to the short-term, medium-term, and long-term outcomes/impacts (see figure 5). There should be at least one evaluation question for each outcome/impact. From each evaluation question, a key indicator or indicators must be identified. These provide the basis for gathering measures to assess whether outcomes/impacts have been achieved. This section discusses the process of setting indicators and the integration of these indicators and related evaluation questions into the logic model.

### Setting Indicators

*What is an Indicator?* Indicators define the data that will be collected in an evaluation and provide evidence to answer the evaluation questions. Indicators should be direct, specific, useful, practical, adequate, and culturally appropriate (Taylor-Powell, Jones, and Henert 2002). We describe in more detail below the different types of indicators.[6]

- *Direct.* Indicators should measure as directly as possible what they are intended to measure. If this is not possible, proxy, or less direct, measures may be used. For example, a direct indicator of MARC program success would be the number and percentage of program participants who complete a graduate degree in a biomedical research field.
- *Specific.* Indicators should be defined in such a way that everyone can have the same understanding of them and understand the data that need to be collected. For example, an indicator of the success of the ADVANCE Program in promoting equitable pay is worded as follows: "Salaries of women faculty in STEM departments compared to those of male faculty in STEM departments (by rank and years in rank)." The wording makes it clear that comparisons of salaries should occur only between male and female faculty of similar ranks and years in rank.
- *Useful.* Indicators should contribute to our understanding of what we are measuring. For example, one indicator of MBRS success would be the number and percentage of participants who obtain employment in a biomedical or behavioral research occupation. Because a main goal of MBRS is to increase the number of underrepresented minorities working in biomedical or behavioral research, this indicator is useful to our understanding of the program focus.
- *Practical.* The cost—in terms of time and money—of collecting data for an indicator should not be more than the utility of the information collected. For example, although it might be interesting to know what math and science courses MARC participants took in high school, the usefulness of this information for the evaluation would not be worth the cost and effort of obtaining and reviewing high school transcripts.

**Figure 5.** LSAMP Program Model with Summative Evaluation Questions and Indicators

| Inputs | Outputs | | Outcomes/Impacts | | |
|---|---|---|---|---|---|
| | **Activities** | **Participants** | **Short-term** | **Medium-term** | **Long-term** |
| Funding from NSF<br><br>Resources from institutions (including in-kind contributions)<br><br>Program research base (tested models, theoretical frameworks, research and evaluation studies, etc.) | Establish partnerships<br><br>Conduct student professional-ization activities<br><br>Improve curriculum<br><br>Develop faculty | Partner institutions<br><br>Student participants<br><br>Faculty participants | Student achievement in STEM<br><br>Student graduation from STEM majors<br><br>Curricular reform<br><br>Improvement of STEM faculty teaching | Student enrollment into a STEM graduate program<br><br>Student attainment of a graduate degree in STEM<br><br>Revisions to institutional policies and practices | Student entry into STEM careers<br><br>Institutionalization of LSAMP practices at participating institutions<br><br>Contributions to scholarly body of work on effective models for increasing minority participation in STEM |

Key Evaluation Questions

| | | |
|---|---|---|
| Are LSAMP participants achieving at a higher rate?<br><br>Are LSAMP participants completeing w/STEM undergraduate degrees?<br><br>Are LSAMP institutions reforming STEM curriculum?<br><br>Has the instruction of STEM faculty at LSAMP institutions improved? | Have LSAMP graduates enrolled in STEM grad programs?<br><br>Have LSAMP graduates completed STEM grad degrees?<br><br>Have LSAMP institutions revised policies and practices to encourage success in STEM? | Have LSAMP graduates entered the STEM workforce?<br><br>Have LSAMP grantees institutionalized program strategies and practices?<br><br>Has LSAMP made a contribution to the research base of effective models for increasing minority participation in STEM? |

Notice how each evaluation question has its own indicator!

Indicators

| | | |
|---|---|---|
| Distribution of GPA of LSAMP participants<br><br>#, % completing STEM undergraduate degrees<br><br>#, % of LSAMP institutions reforming STEM curriculum<br><br>#, % of faculty participating in faculty development activities | #, % of LSAMP graduates enrolled in STEM grad programs<br><br>#, % of LSAMP graduates completing STEM grad degrees<br><br>#, % of LSAMP institutions with revised policies and practices | #, % of LSAMP graduates who enter STEM workforce<br><br>#, % of LSAMP grantees that institutionalize program strategies and practices<br><br>#, % of LSAMP reports and publications in refereed journals and elsewhere about effectiveness of LSAMP model |

*Note:* In this example, the evaluation depicted is a summative evaluation, thus evaluation questions and indicators are only developed for program Outcome/Impacts. If the evaluation also included a *formative* component, the evaluator would also develop questions and indicators for the Inputs and Outputs portions of the logic model.

- *Culturally appropriate.* Indicators need to be relevant within the cultural context. An example of an irrelevant indicator would be the rate of participation of MBRS program enrollees in the International Baccalaureate program in high school.
- *Adequate.* Although there is no correct number or type of indicators, the indicators chosen are a function of what is being measured, the information needed, and the resources available. Often more than one indicator is necessary. Take, for example, the multiple indicators of success for ADVANCE that include number and percentage of women faculty in tenured positions by STEM department, on promotion/tenure committees, awarded tenure or promoted, and several others (all compared with numbers and percentages of men in similar situations).

*Stakeholder Input in Setting Indicators.* Setting indicators through stakeholder input is a complex, time-consuming, and important process. As suggested above, in order for an evaluation to be credible to a variety of stakeholders, there should be agreement among stakeholders regarding the specific indicators that will be used to *prove* whether a program is successful in fulfilling its goals. In soliciting stakeholder input and consensus in determining the indicators used in this project, we used several strategies.

Consensus with different groups of stakeholders can be achieved in various ways. For this project, the process for obtaining input, feedback, and consensus differed by group. We held meetings with the program staff of each organization involved in the project to solicit their input from the program perspective regarding the most appropriate indicators; we also conducted focus groups with project directors/PIs and evaluators for each program to solicit not only input and consensus, but also to explore the feasibility of collecting specific data. Finally, to obtain the input and consensus of expert evaluators from across the nation on indicators and related measures, we conducted an online Delphi exercise with nine expert evaluators.

Indicators identified for each of the programs are listed in appendix B. These indicators, judged to be "very important" or "important" through the process described above, represent stakeholder consensus.

### Integrating the Logic Model with Evaluation Questions and Indicators

Figure 5 illustrates evaluation questions and indicators integrated into a logic model. In the interest of continuity, we once again use the LSAMP Program as an example.

Key evaluation questions developed from the outcomes are linked to the indicators. Measures must then be identified and sources of these measures found. The following section discusses ways of selecting appropriate measures for indicators.

## Measurement Strategies

### Selecting Measures for Indicators

The task of selecting measures for indicators can be difficult. While it is important for reasons of validity to select appropriate measures, issues of time and cost must also be considered. And it is important to remember that sometimes the best measure is the simplest one.

*Self-defined Indicators.* Measures should emerge from indicators identified in the evaluation plan. Sometimes indicators are themselves measures. The short-term indicator shown in figure 5, distribution of GPA of LSAMP participants, is an example of such an indicator. Indicators that can be measured by the number or percentage of entities meeting a quantifiable milestone (for example, number and percentage of LSAMP graduates completing STEM graduate degrees) are in this category of self-defined indicators.

*Performance and Attitudinal Indicators.* Indicators are more difficult to measure when they require a judgment as to whether behavioral, attitudinal, or cognitive change has occurred. These indicators call for measures with a high degree of validity and reliability. Additionally, multiple measures of attitudinal indicators may be required to strengthen the credibility of findings. For example, career interest and attitudes toward science careers are MBRS subindicators for indicator 3, number and percentage of participants who complete a graduate degree in biomedical or behavioral research. Strategies to measure performance or changes in performance (e.g., changes in behavior, knowledge, or skill levels) include the following: selected response tests, constructed response tests, performance simulations, self-report measures, performance indicators, and data (Morris, Fitz-Gibbon, and Lindheim 1991). Measurement approaches to attitudinal change (i.e., changes in affect, feelings, values, or beliefs) include self-report measures, reports of others, sociometric procedures, and records (Henerson, Morris, Fitz-Gibbon 1991).

The following is a description of characteristics of measures for performance and attitudinal indicators.

### *Performance indicators*

- *Selected response tests:* Paper-pencil tests where respondents choose from among various responses. Common formats are multiple choice and true and false.
- *Constructed response tests:* Tests that require examinees to compose an answer to a question by writing a short answer, solving a problem, or delivering an oral presentation. Assessments of writing ability where examinees display their writing ability by composing short essays or paragraphs are examples of this format.
- *Performance simulations:* This type of test requires examinees to perform tasks demonstrating knowledge or skills under controlled conditions similar to real life. An example of this format might be a hands-on scientific problem-solving task.[7]
- *Self-report measures:* This is where respondents provide information about the degree to which behaviors of interest are occurring. For example, high school students might report an increase in participation in math or science after-school activities.
- *Evaluator-developed performance indicators:* A type of measurement that requires the evaluator to develop indicators that demonstrate success in reaching program goals or objectives. For example, increases in student proficiency in math might be measured by the number of problems solved correctly within a specified time.
- *Extant data:* This strategy involves the use of data that might have been collected for other purposes. State assessment results, for example, may give a good indication of student progress in acquiring certain math concepts or skills.

***Attitudinal indicators***

- *Self-report measures:* Participants report directly about their attitudes. These data are often collected through interviews, surveys, polls, questionnaires, and attitude rating scales, logs, journals, or diaries.
- *Reports of others:* Individuals who are in a position to observe participants report about their attitudes. Data may be gathered through interviews, questionnaires, logs, journals, reports, and observation procedures.
- *Sociometric procedures:* Participants in a group report their attitudes toward one another. Peer ratings and social choice techniques are typical methods of data-gathering.
- *Records:* Information on changes in participants' attitudes may be gathered through counselor files or attendance and participation records.

*Sources of Measurement Tools.* Finding performance and attitudinal measures that already exist—as long as they are appropriate measures—has many advantages. By using these measures, evaluators can save time and benefit from others' experiences. Such instruments have probably been field tested and have validity and reliability data.

Already developed performance tests come from three main sources:

- Curriculum materials that have pre- or posttests or curriculum-embedded progress or mastery tests;
- State, district, or funding agency tests that are part of an area-wide assessment program; and
- Tests developed by a test publisher or borrowed from a researcher, professional association, or another source.

Published attitude measures consist mostly of paper and pencil, self-report instruments, many of which may require modification because they were developed for purposes other than program evaluation. Several attitude measures are standardized and available on the commercial test market. "Homemade" instruments developed by a school or school district or copyrighted instruments developed for research are other types of tools available for measuring attitudinal change.

Appendix C contains sources for already published instruments that can be used to collect data on attitudes.

## Demographic Variables

Demographic variables on participants must be collected to answer the key question, "what works for whom in what context?". Here are some common demographic variables as well as issues for cross-project evaluators to consider when using them. It is important to keep in mind that the best versions of these variables are those the funding agency uses to collect and report data. A review of data-collection instruments developed by the funding agency should provide this information. Appendix D provides the demographic variables identified as very important or important for each program by the stakeholders.

*Age*. Age appears to be a straightforward variable; however, there are some pitfalls. For example, asking for date of birth can be problematic because some people will enter the current year rather than their birth year. Providing people with age categories with five or ten year spans (30–39, 30–34, 35–39) is often done, but it limits analysis flexibility including the degree to which age can be used as a dependent variable. For adults, it is generally best just to ask people how old they are.

When the subjects are children, it is often necessary to ask for more than how many years old they are. There can be large development differences between a child who turned five yesterday and one who will turn six tomorrow. For children, it is often useful to ask both their age and the month in which they were born.

*Disability Status*. Disability can be used as a variable at several different levels. Individuals are often asked if they have any disabilities. This question is so broad it is of little value in any analysis. At a more useful level, disabilities can be broken into three levels—physical disabilities, emotional disabilities, and intellectual disabilities, including learning disabilities—and individuals can indicate any areas in which they have disabilities. The most widely used set of definitions comes from the federal law, the Individuals with Disabilities Education Act (IDEA), which breaks disabilities into 13 areas (http://www.nichcy.org/pubs/genresc/gr3.htm#categories):

1. Autism
2. Deaf-blindness
3. Deafness
4. Emotional disturbance
5. Hearing impairment
6. Mental retardation
7. Multiple disabilities
8. Orthopedic impairment
9. Other health impairment
10. Specific learning disability
11. Speech or language impairment
12. Traumatic brain injury
13. Visual impairment including blindness

*Educational Level*. Definitions of educational level are quite standard. The following come from the 2000 Census (http://www.census.gov/prod/2003pubs/c2kbr-24.pdf):

No schooling completed
Nursery school to 4th grade
5th grade or 6th grade
7th grade or 8th grade
9th grade
10th grade
11th grade

12th grade, no diploma
High school graduate—high school diploma or equivalent (for example, GED)
Some college credit, but less than 1 year
1 or more years of college, no degree
Associate degree (for example, AA, AS)
Bachelor's degree (for example, BA, AB, BS)
Master's degree (for example, MA, MS, MEng, MEd, MSW, MBA)
Professional degree (for example, MD, DDS, DVM, LLB, JD)
Doctorate degree (for example, PhD, EdD)

*Job Category.* While there can be many different job categories, the occupational definitions the Census Bureau and the U.S. Equal Employment Opportunity Commission use come the closest to a standard (http://www.eeoc.gov/eeo1survey/jobclassification.html):

- Officials and managers—Occupations requiring administrative and managerial personnel, who set broad policies, exercise responsibility for executing these policies, and direct individual departments or special phases of a firm's operation.
- Professionals—Occupations requiring either college graduation or experience that provides a comparable background.
- Technicians—Occupations requiring a combination of basic scientific knowledge and manual skill that can be obtained through two years of post–high school education, such as is offered in many technical institutes and junior colleges, or through on-the-job training.
- Sales—Occupations engaged wholly or primarily in direct selling.
- Official and clerical—Administrative support occupations, including all clerical-type work regardless of difficulty, where the activities are not predominately manual, though some manual work not directly involved with altering or transporting the products is included.
- Craft workers (skilled)—Manual workers of relatively high level (precision production and repair) having a thorough and comprehensive knowledge of the process involved in their work. They exercise considerable independent judgment and usually have received extensive training.
- Operative (semiskilled)—Workers who operate transportation or materials moving equipment, who operate machine or processing equipment, or who perform other factory-type duties of intermediate skill level that can be mastered in a few weeks and require only limited training.
- Laborers (unskilled)—Handlers, equipment cleaners, helpers, and other workers in manual occupations that generally require no special training and who perform elementary duties that may be learned in a few days and require the application of little or no independent judgment. Farm workers (laborers) are placed here, as well as farming, forestry and fishing occupations not elsewhere covered.
- Service workers—Workers in both protective and nonprotective service occupations.

*Race/ethnicity*. Race/ethnicity is in many ways the most difficult demographic data to collect. One solution is to use the racial/ethnic categories the funding agency uses to collect and report data. Another possible solution is to use the same procedures as the Census Bureau:

> The minimum categories for race are: American Indian or Alaska Native; Asian; Black or African American; Native Hawaiian or Other Pacific Islander; White and Some Other Race. Respondents can select one or more races when they self-identify. There are also two minimum categories for ethnicity: Hispanic or Latino and Not Hispanic or Latino. Hispanics and Latinos may be of any race. (http://www.census.gov/population/www/socdemo/race/racefactcb.html)

Alternatively, members of the target audiences can be asked which racial category titles are most appropriate for their population. Asian/Pacific Islander is often used as a category. However, since Pacific Islanders are considered an underrepresented minority and Asians are not, it is better to separate them into two categories. “Other” or “some other race” should be used either minimally or not at all since such a category is difficult to interpret. Individuals who indicate their race/ethnicity as “other” also cannot be included as underrepresented minorities.

*Sex/Gender*. Sex/gender is a straightforward variable. The only major decision is if “sex” or “gender” should be used as the variable name. At times gender is used rather than sex so there can be no confusion with questions related to sexuality.

*Socioeconomic Status (SES)*. There are a number of different types of data that can be used to determine SES, none of which are perfect. Public school students’ eligibility for free or reduced lunch is often used as a measure of low-income status. For all age levels, in combination with census data, zip codes can be used as a measure of SES. More information about the use of census data can be found at http://www.census.gov/main/www/access.html.

Family income is also used as an SES measure. Since people tend to know only their approximate family income, it is usually more effective to provide people with income categories with five or ten thousand dollar spans ($30,000–$39,999 or $30,000–$34,999; $35,000–$39,999) than to ask for an exact amount.

Education level and job category can also indicate SES. For children, parents’ education levels and job categories are used; for adults, their own and their spouse’s or partner’s are used. A sample procedure, the Barratt Simplified Measure of Social Status (BSMSS) has been developed for measuring SES and can be found at http://wbarratt.indstate.edu/socialclass/Barratt_Simplifed_Measure_of_Social_Status.pdf.

## Conclusion

As the first of a two-volume guide, “Designing a Cross-Project Evaluation” lays out the process for the design of a program evaluation of multiproject programs. A second volume, “Collecting and Using Data in Cross-Project Evaluations,” follows and provides guidance on the collection, reporting, and use of evaluation data. Together, these guides offer a wealth of information and practical advice that will enable organizations to develop their capacity to conduct cross-project evaluations.

## Notes

1. The underrepresented groups include women, African Americans, Latinos, and American Indians.
2. The authors refer those seeking information on routine evaluation tasks to the NSF "User Friendly Guides" listed in the references.
3. There are situations in which it is helpful to involve stakeholders at this stage, particularly when they are especially knowledgeable about program goals and objectives.
4. In cross-project evaluation, the ultimate beneficiaries of the projects are rarely considered stakeholders.
5. The section on constructing a logic model is based on "Enhancing Program Performance with Logic Models" (Taylor-Powell, Jones, and Henert 2002).
6. This discussion of indicators draws from Taylor-Powell, Jones, and Henert (2002).
7. The use of rubrics or checklists to guide the collection of measurement data on this type of indicator is helpful.

## References


Henerson, M. E., L. L. Morris, and C. T. Fitz-Gibbon. 1991. *How to Measure Attitudes.* Newbury Park: Sage Publications.

Morris, L. L., C. T. Fitz-Gibbon, and E. Lindheim. 1991. *How to Measure Performance and Use Tests.* Newbury Park: Sage Publications.

Stecher, B. M., and Davis, W. A. 1990. *How to Focus an Evaluation.* Newbury Park: Sage Publications.

Stockdill, S. H., M. Baizerman, and D. W. Compton, D. W. 2002. "Toward a Definition of the ECB Process: A Conversation with the ECB Literature." *New Directions in Evaluation* 93: 1–25.

Taylor-Powell, E., L. Jones, and E. Henert. 2002. "Enhancing Program Performance with Logic Models." University of Wisconsin-Extension web site: http://www1.uwex.edu/ces/lmcourse/.

## Appendix A. Program Descriptions

The following section describes the programs chosen from each of the participating organizations.

### National Science Foundation (NSF)

*ADVANCE (Institutional Transformation Award):* Established in 2001 as part of the overall ADVANCE goal of increasing the representation and advancement of women in academic science and engineering careers, the Institutional Transformation Award allows institutions to define and implement approaches to increase the participation and advancement of women faculty members into senior and leadership ranks of science and engineering.

*Alliances for Graduate Education and the Professoriate (AGEP):* This program, established in the mid-1990s, seeks to increase significantly the number of African-American, Hispanic, and American Indian students receiving doctoral degrees in all disciplines funded by NSF. AGEP programs develop and implement innovative models for recruiting, mentoring, and retaining minority students in doctoral programs as well as effective strategies for identifying and supporting underrepresented minorities who wish to pursue academic careers.

*Louis Stokes Alliances for Minority Participation (LSAMP):* LSAMP, established in 1991, is a comprehensive, multidisciplinary, undergraduate program to increase substantially the quantity and quality of URM students completing degrees in STEM and continuing on to graduate programs in STEM fields. The program requires each awardee to establish meaningful partnerships among academic institutions, government agencies and laboratories, industry, and professional organizations.

### National Institutes of Health (NIH)

*Minority Access to Research Careers: Undergraduate Student Training in Academic Research (MARC: U*STAR):* Established in 1977, this program has as its goal increasing the participation of URMs in biomedical research and health care professions. Additional goals are to increase the number and capabilities of scientists from underrepresented minority groups who are engaged in biomedical research and to strengthen science curricula and student research opportunities at institutions with substantial minority enrollments to prepare them for research careers. All projects provide undergraduate students with a summer research experience at a research-intensive institution, academic-year research opportunities, and other educational experiences.

*Minority Biomedical Research Support (MBRS)—Initiative for Minority Student Development (IMSD):* The goal of MBRS, established in 1993, is to develop biomedical and behavioral research scientists who are also URMs. As a subprogram of MBRS, IMSD encourages the development and expansion of innovative programs to improve the academic and research competitiveness of URMs at the undergraduate, graduate, and postdoctoral levels.

## GE Foundation

*Math Excellence Initiative:* The purpose of this program, established in 2001, is to strengthen and expand the pipeline of URMs and women in engineering, information technology, and quantitative business disciplines (finance, accounting, and economics). Grants support comprehensive strategies that have a long-term, sustained impact on strengthening students' math and quantitative problem-solving skills; increasing student knowledge and interest in math and quantitative careers; and increasing college-level recruitment and retention in these fields. Individual projects target elementary, middle, and high school students and teachers, and college students and professors.

# Appendix B. Program Indicators

## ADVANCE-Important Indicators

**Indicator and Additional Measures**

1. Number and percentage of women faculty in tenured positions by STEM department (compared to men)
2. Number and percentage of women accepted into faculty positions in STEM departments (by faculty rank and tenure)
3. Number and percentage of women STEM faculty awarded tenure/promoted (compared to men)
4. Salaries of women faculty in STEM departments vs. male faculty in STEM departments (by rank and years in rank)

   Years in rank
   Faculty salaries

5. Changes in hiring procedures to advance equity
6. Changes in promotion and tenure procedures to advance equity
7. Years in rank of women in faculty positions in the institution (compared to men)
8. Number and percentage of women on promotion/tenure committees (compared to men)
9. Number and percentage of women in high administrative positions
10. Number and percentage of women applying for faculty positions in STEM departments by faculty rank (compared to men)
11. Changes in faculty evaluation to advance equity
12. Establishment of, or increase in, "family friendly" policies toward faculty and graduate students (e.g., availability of day care, flexibility in time schedule for tenure/degree attainment)
13. Institutionalization of recruitment strategies that attract female applicants

## AGEP-Important Indicators

**Indicator and Additional Measures**

1. Number and percentage of underrepresented minorities who enroll in graduate degree programs in STEM

   Graduation with STEM graduate degree (MS, PhD)
   Department/Major
   Degree sought
   Current standing/progress towards degree
   Awards received
   Type of institution (Carnegie classification)

2. Number and percentage of underrepresented minorities who graduate with a doctoral degree in STEM

Department/Major
Type of institution (Carnegie classification)

3. Number and percentage of underrepresented minorities who obtain faculty positions in STEM

   Highest STEM degree attained
   Type of employer
   Major in highest STEM degree
   Faculty rank
   Tenure status

4. Changes in graduate student funding policies to be more reflective of individual student needs

## LSAMP-Important Indicators

**Indicator and Additional Measures**

1. Number and percentage of participants who complete a baccalaureate degree in a STEM field Major

   Type of institution (Carnegie classification)
   GPA
   Participation in program activities (by activity type)
   Presentations and publications

2. Number and percentage of participants who enter a graduate program in STEM

   Department/Major
   Degree Sought
   Graduation with STEM graduate degree (MS, PhD)
   Type of institution (Carnegie classification)

3. Number and percentage of participants who complete a graduate degree in STEM

   Degree attained
   Type of institution (Carnegie classification)

4. Number and percentage of participants who obtain employment in a STEM occupation

   Major in highest STEM degree
   Highest STEM degree attained
   Type of employer
   Position

5. Changes in the curriculum at the department level to advance equity

## MARC U*STAR-Important Indicators

**Indicator and Additional Measures**

1. Number and percentage of participants and underrepresented non-participants who enter a STEM undergraduate major leading to a biomedical science research career

    Major
    Degree sought
    Graduation with a STEM Baccalaureate degree (follow-up measure)
    High School GPA
    Participation in program activities (by activity type)
    Student research skills
    Confidence in math
    Attitudes towards science and math careers
    Standardized test scores, e.g., SAT, ACT

2. Number and percentage of participants and underrepresented non-participants who receive a baccalaureate degree in STEM that can lead to a biomedical science research career

    Major
    Degree attained
    Participation in program activities (by activity type)
    GPA
    Student research skills
    Presentations and publications (# and type)
    Attitudes towards science and math careers
    GRE test scores

3. Number and percentage of participants and underrepresented non-participants who enter a PhD program in a biomedical or behavioral science field

    Department/Major
    Current standing/progress towards degree
    GPA (Undergraduate)
    Attitudes towards science and math careers
    Presentations and publications (# and type)
    Type of institution (Carnegie classification)
    GRE test scores

4. Number and percentage of participants who complete a PhD in a biomedical or behavioral research field

    Department/Major
    Degree attained
    Type of institution (Carnegie classification)

Presentations and publications (# and type)
Career interests/intent
Time to degree

5. Number and percentage of participants who obtain postdoctoral training or research employment in a biomedical science research field

    Major in highest STEM degree
    Highest STEM degree obtained
    Type of employer
    Position
    Career focus

6. Number and percentage of students at institutions with substantial minority enrollments receiving meaningful research opportunities

    Major
    Graduation with a STEM Baccalaureate degree
    Current standing/progress towards degree
    Courses being taken
    Courses grades
    GPA
    Student research skills
    Presentations and publications (# and type)
    Attitudes towards science and math careers
    Initial career interest
    Standardized test scores, e.g., SAT, ACT

7. Increase in the quality of the science curriculum at institutions with substantial minority enrollments
8. Number and percentage of participants who take the GRE
9. Increase in the number and level of required math and quantitative science courses in institutions with substantial minority enrollments

## Math Excellence-Important Indicators

**Indicator and Additional Measures**

1. Number and percentage of participants who show increases in knowledge about math and quantitative careers

    Courses being taken
    Course grades
    Confidence in math
    GPA
    Attitudes towards math and math careers
    Initial career interests

2. Number and percentage of participants who successfully complete AP high school math courses

    Courses being taken
    Confidence in math
    Scores on AP exams
    Course grades
    GPA
    Initial career interests
    Attitudes towards math and math careers
    Standardized test scores (changes over time)

3. Number and percentage of participants who show an increased interest in math and quantitative careers

    Course being taken
    Course grades
    GPA
    Confidence in math
    Attitudes towards math and math careers
    Initial career interests

4. Number and percentage of participants who enter college majors in quantitative disciplines (i.e. math, physical sciences, engineering, finance, accounting)

    Major
    Degree sought
    Current standing/progress towards degree
    Courses being taken
    Course grades
    GPA
    Confidence in math
    Attitudes towards math and math careers
    Initial career interests
    Standardized test scores, e.g., SAT, ACT

5. Number and percentage of participants who graduate with a baccalaureate degree in a quantitative discipline field

    Major
    Degree attained
    Courses being taken
    Course grades
    GPA
    Confidence in math
    Attitudes towards math and math careers

6. Number and percentage of teachers in collaborating school systems who received training from the project
7. Number and type of collaborations between school systems and colleges/universities
8. Number and percentage of participants successfully completing Algebra I in the 8th grade Courses being taken

    Course grades
    GPA
    Confidence in math
    Attitudes towards math and math careers
    Initial career interests

## MBRS-Important Indicators

**Indicator and Additional Measures**

1. Number and percentage of participants who graduate with a baccalaureate degree in biomedical or behavioral research

    Major
    Degree attained
    Course grades
    GPA
    Presentations and publications (# and type)
    Attitudes towards science careers

2. Number and percentage of participants who enter a graduate program in biomedical or behavioral research

    Department/Major
    Degree sought
    Graduation with STEM graduate degree (MS, PhD)
    Type of institution (Carnegie classification)
    Current standing and progress toward degree
    Courses being taken
    GPA
    Student research skills
    Presentations and publications (# and type)
    Attitudes towards science careers
    Initial career interest
    Standardized test scores, e.g., GRE

3. Number and percentage of participants who complete a graduate degree in biomedical or behavioral research

    Department/Major
    Degree attained

Type of institution (Carnegie classification)
GPA
Presentations and publications (# and type)
Attitudes towards science careers
Career interests
Standardized test scores, e.g., GRE

4. Number and percentage of participants who obtain employment in a biomedical or behavioral research occupation

Major in highest STEM degree
Highest STEM degree attained
Type of employer
Position
Presentations and publications (# and type)
Attitudes towards science careers

# Appendix C. Published Instruments for Collecting Attitudinal Data

## Examples of Instruments that Measure Student Attitudes

| Instrument | Grade Level | Author | Pub. Date | Source |
|---|---|---|---|---|
| General Academic Motivation | | | | |
| Academic Motivation Scale | High School & Adult | Robert J. Vallerand, et al | 1993 | *Educational and Psychological Measurement, 53,* 159–172. |
| Affective Perception Inventory, Advanced Level | High School | Soares & Sorares | 1979 | Soares Associates; 111 Teeter Rock Rd., Trumbull, CT 06611 |
| Estes Attitude Scales: Secondary Form | 7–12 | Estes, Estes, Richards, and Roettger | 1981 | PRO-ED; 5341 Industrial Oaks Boulevard, Austin, TX 78735 |
| Intellectual Achievement Responsibility Questionnaire | 1–12 | Crandall, Katkovsky, and Crandall | 1965 | *Child Development, 36,* 91–109. |
| Inventory of School Motivation (ISM) | High School | McInerney & Swisher | 1995 | *Journal of American Indian Education, 34*(3), 28–51. |
| Motivated Strategies for Learning Questionnaire (MSLQ) | Middle School | Pintrich and DeGroot | 1990 | *Journal of Educational Psychology, 82,* 33–40. |
| School Attitude Measure | 1–12 | ACT | 1989 | Standardized test, ACT |
| Attitudes Towards Math | | | | |
| Aiken's Enjoyment of Mathematics and Value of Mathematics Scales | High School & Undergraduate | Aiken | 1983 | *Educational and Psychological Measurement, 43*(4), 1247–53. |
| Fennema-Sherman Mathematics Attitudes Scales | 9–12 | Fennema and Sherman | 1976 | *Journal of Research in Mathematics Education, 7*(5), 324–326. |
| Indiana Mathematical Belief Scales | High School & Undergraduate | Kloosterman & Stage | 1992 | *School Science and Mathematics, 92*(3), 109–15. |
| Math Anxiety Questionnaire | 6–12 | Wigfield and Meese | 1988 | Department of Human Development, Institute for Child Study, University of Maryland, College Park |
| Mathematical Self-Concept Scale | Undergraduate | Gourgey | 1982 | ERIC Document Reproduction Service (ED 223 702; 18 pages) |
| Mathematics Self-Efficacy Scale | High School & Adult | Betz & Hackett | 1993 | Mind Garden, P.O. Box 60669, Palo Alto, CA 94306 |
| National Assessment of Educational Progress, Released Exercises: Mathematics. | 4, 8, 12 | | 1983 | National Assessment of Educational Progress; Box 2923; Princeton, NJ 0854 |
| Revised Math Attitude Scale | Undergraduate | Aiken | 1963 | *The Journal of Educational Research, 56*(9), 476–80 |
| Students' Beliefs about Mathematics | 12-Jan | Spangler | 1992 | Arithmetic Teacher, 40(3), 148–52. |

*(continued)*

## Appendix C. *Continued*

| Instrument | Grade Level | Author | Pub. Date | Source |
|---|---|---|---|---|
| Who and Mathematics | 9–12 | Leder and Forgasz | 2002 | *Two new instruments to probe attitudes about gender and mathematics.* (ERIC Document Reproduction Service No. ED 463312). |
| Attitudes Towards Science | | | | |
| Attitudes Toward Science Inventory | Middle and High School | Gogolin and Swartz | 1992 | *Journal of Research in Science Teaching, 29,* 487–504. |
| Attitude Towards Science (ATS) | 4–5 | Klausmeier, DiLuzio, Brumet | 1976 | Wisconsin Research and Development Center for Cognitive Learning |
| Attitudes Toward Science in School Assessment | 7–8 | ETS | 1988 | ETS TestLink |
| Modified Attitudes Toward Science Inventory | Middle School urban students | Weinberg and Steele | 2000 | *Journal of Women and Minorities in Science and Engineering, 6*(1) |
| Science Anxiety Questionnaire | Undergraduate | Mallow | 1994 | *Journal of Science Education and Technology, 3*(4), 227–38. |
| Science Attitude Scale for Middle School Students | 5–8 | Misiti, Shrigley, Hanson | 1991 | *Science Education, 75*(5), 525–540. |
| Test of Science Related Attitudes (TOSRA) | High School | Fraser | 1981 | Australian Council for Educational Research |
| Views on Science-Technology-Society | High School | Aikenhead, Ryan and Fleming | 1989 | *Views on science-technology-society* (Saskatoon, Sasatchewan: Department of Curriculum Studies) |
| Attitudes Towards STEM Study and Careers | | | | |
| Longitudinal Survey of American Youth (LSAY)* | Middle School to Adult | | 1987 | Michigan State University www.lsay.org |
| High School and Beyond* | High School & Adult | | 1980 | National Center of Education Statistics |
| Image of Science and Scientists Scale | High School | Krajkovich, Joseph G | 1978 | ETS TestLink |
| Interest in Scientific Activities | High School | Gardner & Tamir | 1989 | *Journal of Research in Science Teaching, 26*(5), 409–23. |
| Schwirian Science Support Scale | Undergraduate and Adult | Schwirian | 1968 | *Science Education, 52*(2), 172–79. |
| Science Research Temperament Scale | High School & Undergraduate | Kosinar | 1955 | Psychometric Affiliates |
| Test of Perceptions of Scientists and Self | High School | Mackay & White | 1976 | ETS TestLink |
| Women in Science Scale | Middle & High School | Erb & Smith | 1983 | *Journal of Research in Science Teaching, 21*(4), 391–97. |

*Note:* A rich source for published testing instruments is ETS TestLink (http://sydneyplus.ets.org). This list was compiled using a search for "attitude measures."

## Appendix D. Demographic Variables by Program

### Demographic Indicators Deemed "Important" or "Very Important," by program

**Program and Indicators**

ADVANCE
- Sex
- Department/Administrative Office
- Faculty Rank
- Tenure Status
- Race/Ethnicity

AGEP
- Race/Ethnicity
- Sex
- First generation college

LSAMP
- Race/Ethnicity
- Sex
- First generation college

MARC U*STAR
- Race/Ethnicity
- Sex
- First generation college

MBRS
- Race/Ethnicity
- Sex

The Urban Institute

2100 M Street, NW
Washington, DC 20037

Phone: 202.833.7200
Fax: 202.429.0687
E-mail: paffairs@ui.urban.org
http://www.urban.org