

Running head: COMMON CORE INDICATORS

Common Core Indicators for Describing Alliance Programs

Tom McKlin

The Findings Group, LLC

Abstract

Under growing scrutiny among policymakers, many NSF program officers ask evaluators to design, collect, and report on a set of indicators common across a portfolio of programs. This presentation specifically addresses the issues of establishing, reporting, and ultimately using common, core indicators. This discussion draws on three sources:

- The experience of evaluating multiple National Science Foundation (NSF) alliance programs in Research in Disabilities Education (RDE) and Broadening Participation in Computing (BPC).
- The experience over the past four years of working with a small group of alliance evaluators to define common indicators and to report on those indicators.
- Recent publications guiding alliance evaluators on establishing common indicators, namely the Framework for Evaluating Impacts of Broadening Participation Projects (Clewell & Fortenberry, 2009) and the Framework for Evaluating Impacts of Informal Science Education Projects (Friedman, 2008)).

Based on the work of creating common, core indicators and studying those created under other NSF programs, shared elements among common, core indicators emerge. Common, core indicators focus largely on counting the number of participants in program activities, tracking students through transitions (e.g. high school to college or college to graduate school), measuring changes in affective characteristics of participants, and building the capacity of funded organizations like colleges and public schools.

This work also reveals myths among evaluators surrounding NSF's treatment of annual reports and data. NSF does not mandate a specific set of metrics for programs. They request that programs identify broader impacts, but when looking across alliances, these rarely, if ever, align and do not require that proposers focus on diversity (gender, race/ethnicity, and ability). Many evaluators mistakenly think that a group at NSF (or other agencies) is synthesizing reports within a program or directorate. This activity is not happening. Some evaluation findings are rolled into various reports, and these reports look more like salad than soup in that they are presented as a collection rather than a synthesis.

Finally, this presentation invites evaluators to identify areas of immediate action. For example, one of the most challenging aspects of this work is the need to track students through transitions. This is notoriously difficult and requires the best thinking among the evaluation community to reliably and efficiently track participants. Second, evaluators may request to see how program officers and others use the common core indicators. We often report to a program officer yet have little understanding of what becomes of the information once it is sent. Seeing first-hand how the data are used aides the evaluation community in building more effective indicators and reporting mechanisms. It also signals that the request is valid instead of a futile and expensive exercise in report generation.

Common Core Indicators for Describing Alliance Programs

As budgets for federal agencies like the National Science Foundation (NSF) and the National Institutes for Health (NIH) draw increased scrutiny, agencies are under greater pressure to provide evidence that their programs are working. In September, 2011, the Senate voted to cut NSF's budget by 2.4% (\$162 million) and \$626 million across all agencies under the Subcommittee on Commerce, Justice, Science, and Related Agencies. The chair of the Senate subcommittee, Senator Barbara Mikuski, said, "for the first time as chair, I've eliminated programs" (Mervis, September 11, 2011). This past May, the House of Representatives discussed and ultimately rejected a bill that would cut NSF's budget by \$1.2 billion (Jones, May 23, 2012).

Relatedly, many have asked that agencies increase and improve their data collection, analysis, and reporting practices. Given the task to assess programs aimed at improving America's competitiveness in STEM, the Academic Competitiveness Council wrote, "The ACC urges Congress to give careful consideration to the results of program and project impact, and to hold support for funding increases in abeyance until it is determined that programs have the capacity to rigorously evaluate the effectiveness of their activities" (U. S. Department of Education, 2007). Not only does the ACC call for rigorous evaluation and withholding funds in the absence of rigorous evaluation, they also recommend common indicators: "Agencies will establish common metrics and collect common data elements among all projects to enable comparative assessments that will yield information about best practices." Clewell and Fortenberry (2009) make a similar claim.

There is a need for reliable, consistent, and more detailed data from PIs about students, postdoctoral researchers, and staff supported by their grants.... These reports should use a common set of questions, either across all programs or across program types.... (Clewell & Fortenberry, 2009).

While these claims for common metrics occurred before the latest round of budget scrutiny at NSF, one can only imagine that increased pressure to justify program expenditure would lead to enacting past recommendations to rigorously evaluate at the program level.

Definitions

This paper refers to both "program evaluation" and "project evaluation" and uses the phrase "program evaluation" differently than most evaluators do. Here, program evaluation is an evaluation of an agency (e.g. NSF, NIH, etc.) program like Broadening Participation in Computing's Alliance Programs that is comprised of many funded projects. Each project has its own evaluation. Many evaluators use the phrase "program evaluation" to refer to the evaluation of a single project funded under an agency program.

One response to the need for program-level evaluation is to develop a set of common metrics, questions, or instruments to be used across projects. This type of multi-site evaluation is commonly done by an external research and evaluation firm and is often different than the evaluation organizations that conduct the individual project evaluations. Occasionally, this set of common, core indicators grows organically from among the evaluators and PIs working on projects within one agency program. These indicators are

“common” across all projects in the program and “core” in that they aim to measure what is central, similar, and focused on the program’s primary purpose. They are also “core” in the sense that they do not propose to measure all aspects of each project. In its perfect form, the project evaluation plan would wrap neatly around and compliment the common, core indicators, adding formative feedback for mid-course correction and summative reflection for future projects.

Developing the Indicators

NSF’s Broadening Participation in Computing program enlisted the American Association for the Advancement of Science (AAAS) to coordinate the program’s alliance evaluators to develop, collect, and champion a set common, core indicators that would describe the merit and worth of its alliance programs. This approach is beneficial because it provides an early indication for the type of data that are feasible to collect, makes the evaluators aware of instruments and techniques used across projects, enlists the support of the evaluators, provides the evaluators with an early idea of the indicators to be collected. Engaging the project evaluators in this way makes the larger program evaluation seem less imposing and potentially more useful. This approach also allows the evaluators and PIs to discuss and agree upon the most central and critical measurement elements and encourages cross-project evaluation capacity-building among the evaluators. Evaluators drew on the expertise and ideas of other evaluators and PIs.

Taking the opposite approach, asking an external organization to develop common, core indicators in a without project evaluator support, is potentially detrimental. The project evaluators and PIs will comply with the multi-site program evaluators, they have to, but they will do so in the spirit of compliance and monitoring, not in the spirit of providing program decision-makers with a viable justification for their expenditures.

Numerous teams within the National Science Foundation alone have tackled the development of common, core indicators. Here, we look at two sets of frameworks for common core indicators: the Framework for Evaluating Impacts of Broadening Participation Projects (Clewell & Fortenberry, 2009); and the Framework for Evaluating Impacts of Informal Science Education Projects (Friedman, 2008). Added to this is the set of common, core indicators that the Broadening Participation in Computing evaluators compiled. Tables 1 and 2 show all three sets of common, core indicators and separates them into indicators related to project participants and those related to organizations.

Notice that both Broadening Participation and Informal Science focus on the individual participant; however, they do so in very different ways. Broadening Participation is primarily interested in increasing the number and diversity (hence, “broadening participation”) of participants progressing through transition points (e.g. college to graduate school; graduate school to the professoriate). Looking at the outcomes of the logic model for a typical program (see Figure 1), we may categorize these as observable outcomes. For example, we can observe someone entering an undergraduate program or receiving a doctorate in science and engineering. In contrast, Informal Science is primarily interested in measuring the internal characteristics of their participants. While they are

also interested in the number participating in their projects, their indicators demonstrate that they are ultimately interested in making qualitative, internal changes among participants. For example, Informal Science seeks data related to knowledge, attitudes, and skills. Informal science is also interested in observable behaviors (such as making healthy food choices, conserving energy, and limiting water usage) that typically accompany changes in knowledge and attitudes. Looking back at the typical program logic model, Informal Science primarily seeks *internal* participant outcomes.

General Program Logic Model

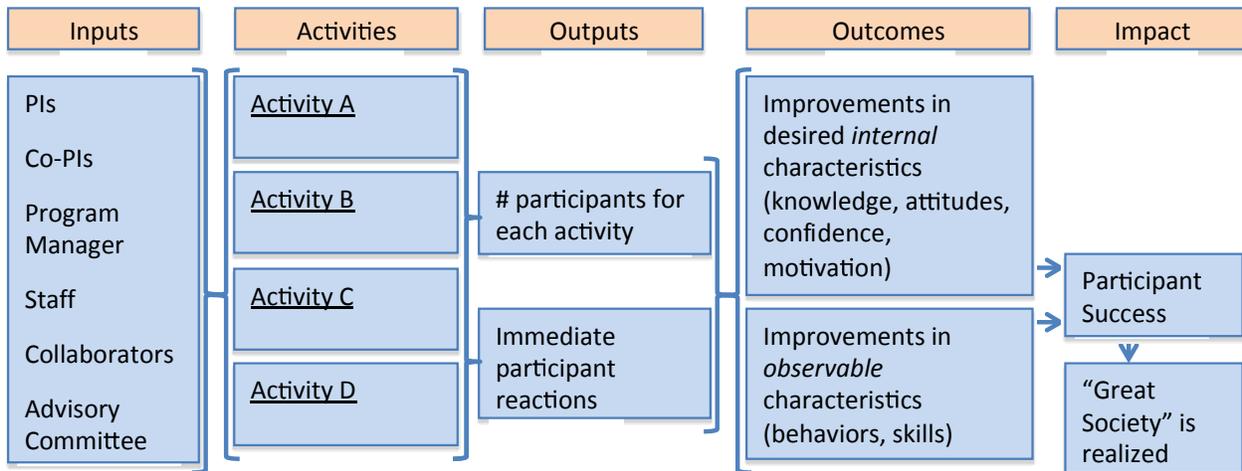


Figure 1

The Broadening Participation in Computing evaluators acknowledged the importance of both sets of indicators: observable transitions and changes in internal characteristics that often accompany those transitions. At a more basic level, though, these evaluators felt it important to report on the types of activities occurring across programs. For example, many programs provide workshops and summer camps for students along with professional learning opportunities for teachers. Like Informal Science, the evaluators included internal characteristics and included “intention to persist” as an internal characteristic that informs transitional behavior. And like Broadening Participation, the evaluators included observable indicators such as progressing through transition points.

The Informal Science indicators primarily focus on the individual, while the Broadening Participation indicators extend beyond the individual to organizations in an effort to sustain the effects of the program on individuals who may be influenced by the organization after the program has ended. Broadening Participation focuses on three sustainability areas designed to increase the number and diversity of individuals influenced by the organization: institutional policy; increased research and teaching capacity; and increased collaboration (see Table 2).

Similarly, the Broadening Participation in Computing evaluators sought to measure increased capacity in three ways. First, they sought to measure increased capacity in

organizations directly influenced by the program such as policy changes within university departments or professional learning for K-12 teachers in partner schools. Second, they sought to measure the effect the program had on organizations not directly supported by the program such as the dissemination of promising practices to other organizations or establishing statewide policies (like articulation agreements) affecting multiple universities. Finally, the Broadening Participation in Computing evaluators sought to measure community-building as a cross-cutting indicator affecting both individuals and organizations. They drew from the work of Wenger et al. (2011) which outlines a framework for studying networks and communities and acknowledged that many of the Broadening Participation in Computing projects intentionally sought to expand networks and build communities. Here again is another area that an external evaluation organization may have missed. The evaluators only came to realize the pervasiveness of intentional community-building through extensive conversations with PIs and evaluators around the development of common, core indicators. Perhaps as a benefit of the organic development of the common, core indicators, the Broadening Participation in Computing evaluators addressed many of the Broadening Participation areas (namely, institutional policy and teaching capacity) while taking a broader view of organizational improvement and community-building.

Table 1. Indicators Related to Individuals

Outcome Type	Program		
	Broadening Participation	Informal Science Education	Broadening Participation in Computing
External (Observable)	<ol style="list-style-type: none"> 1. Individual-focused programs (goals): Increase the number of individuals: <ol style="list-style-type: none"> a. Entering undergraduate majors in S&E. b. Receiving a baccalaureate degree in a S&E field c. Entering into a graduate S&E program. d. Receiving a doctorate in S&E. e. Entering the professoriate/work-force in S&E. f. Increased progress and advancement of faculty in S&E academe or research. 		<ol style="list-style-type: none"> 1. Individual Participation and Outcomes: <ol style="list-style-type: none"> a. Activities (type, duration, level) b. Measures of Internal Characteristics (intention to persist, engagement, confidence, knowledge/skills) c. Observable Indicators/Changes (Progress: transitioning from one academic level to the next)
Internal		<ol style="list-style-type: none"> 1. Awareness, knowledge or understanding 2. Engagement or interest 3. Attitude 4. Behavior 5. Skills 6. Other (project-specific and unintended outcomes) 	

Note: Broadening Participation indicators are primarily external and observable while the Informal Science indicators are primarily internal and related to knowledge or affect.

Table 2. Indicators Related to Institutions

Program	
Broadening Participation	Broadening Participation in Computing
2. Institution-focused programs (goals) <ol style="list-style-type: none"> a. Encourage equitable institutional policies and practices in post-secondary STEM departments. b. Increase research capability and teaching effectiveness in S&E disciplines. c. Encourage collaboration of MSIs with other entities to enhance effectiveness. 	2. Organizational Capacity <ol style="list-style-type: none"> a. Number and types of organization impacted b. Type of Impact (sustain or institutionalize activities, policy change, train/develop skills and knowledge, generate/disseminate tools, expand stakeholder awareness) c. Measurement: description of how the change is measured d. Populations impacted 3. Alliance Impact: effect of the alliance on external organizations <ol style="list-style-type: none"> a. Type of Impact b. Relationship with the alliance c. Description of the change over time Crosscutting Indicator: Building Community <ol style="list-style-type: none"> a. Participation in community or development of network b. Value of participation in community c. Intended value of participation

Note: Informal Science does not formally present organizational indicators.

Reporting Common Core Indicators

During development of the common, core indicators, it became increasingly important to test the metrics by asking project evaluators to report on the common, core indicators. This revealed portions of the indicators that were confusing, impossible to collect or beyond the scope of existing evaluation efforts. Figures 2 and 3 provide examples of reporting elements surrounding the first common, core indicator: Individual Participation. Here, the evaluators sought to describe not only the number of participants but also the dosage. Dosage is broadly categorized as “touched” which means the participants received less than a day of program activity, “limited engagement” meaning the participants received approximately one day of program activity, and “deeper engagement” which means that the participants received more than a day of program activity.

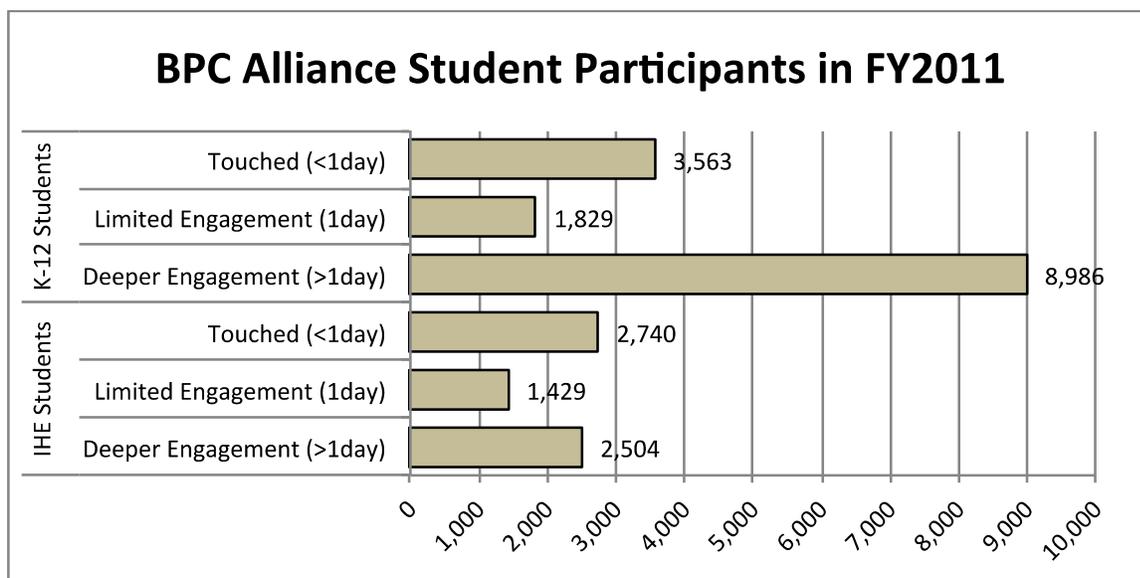


Figure 2

The Informal Science framework recommends measuring internal characteristics of participants (knowledge, engagement, etc.), and many of the BPC programs also measured internal characteristics. However, the programs all did so in vastly different ways. At the programmatic level, the evaluators are primarily interested in whether the program generated evidence of statistically significant gains in internal characteristics. The program evaluators broadly asked this question across all projects in the program and derived the percentage of students experiencing the results of significant program effect (see Figure 3).

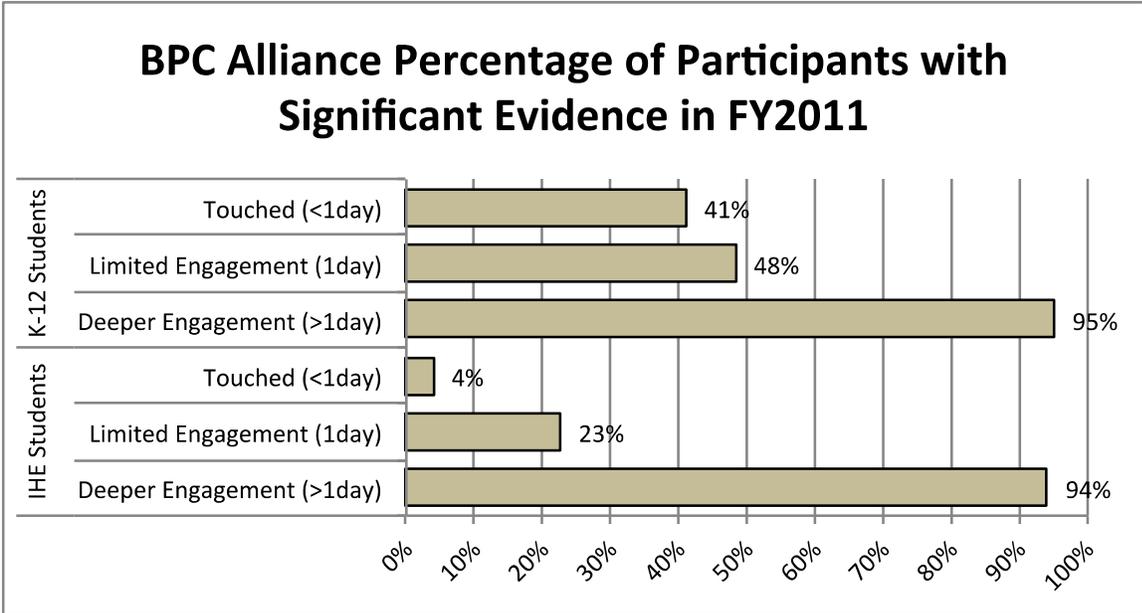


Figure 3

The second common, core indicator examines the effect the program has on organizations directly influenced by each project. Figures 4 and 5 are examples of how these data are consolidated across projects. Figure 4 shows the number and types of organizations affected by all projects in the program, and Figure 5 describes how these organizations were affected.

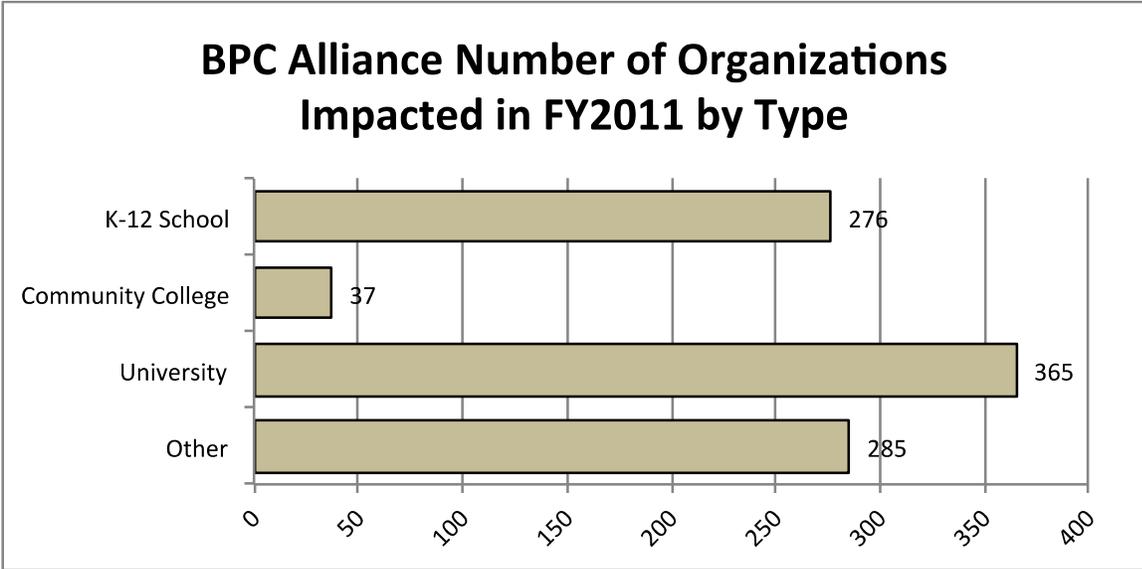


Figure 4

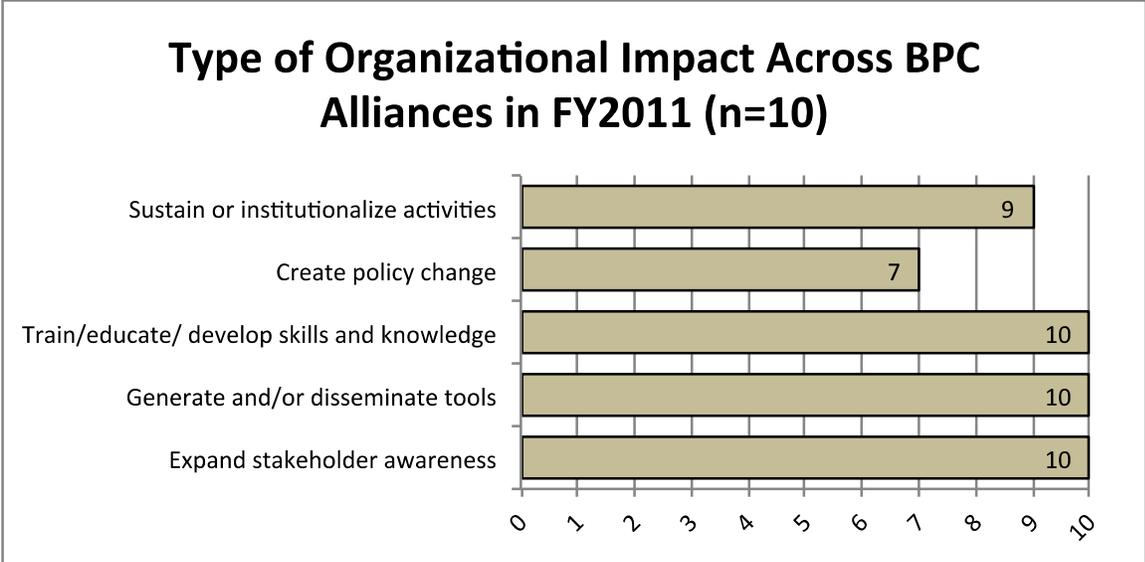


Figure 5

The third common, core indicator examines the effect of the program on organizations that are not formally identified as supported partners in a specific project. Figures 6 and 7 exemplify how the evaluators reported these data. Figure 6 shows the types of organizations affected and the number of projects (or alliances) affecting them. Figure 7 shows how the external organizations were broadly affected by the program.

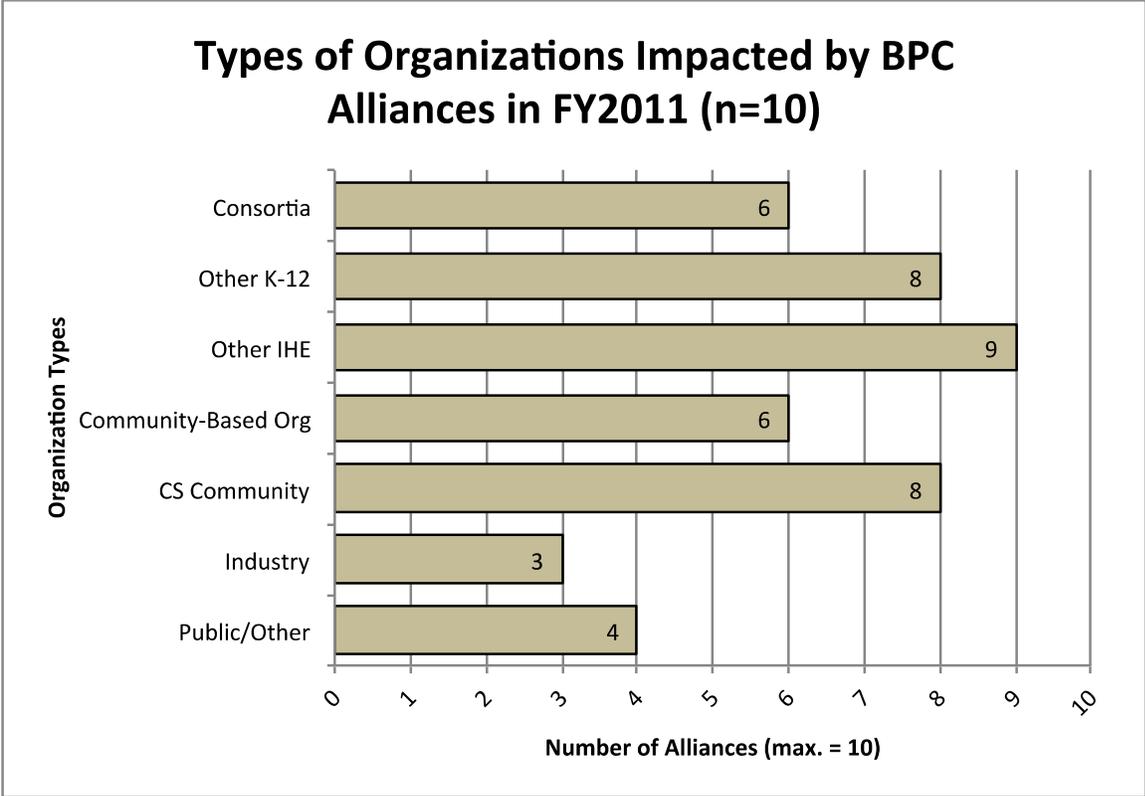


Figure 6

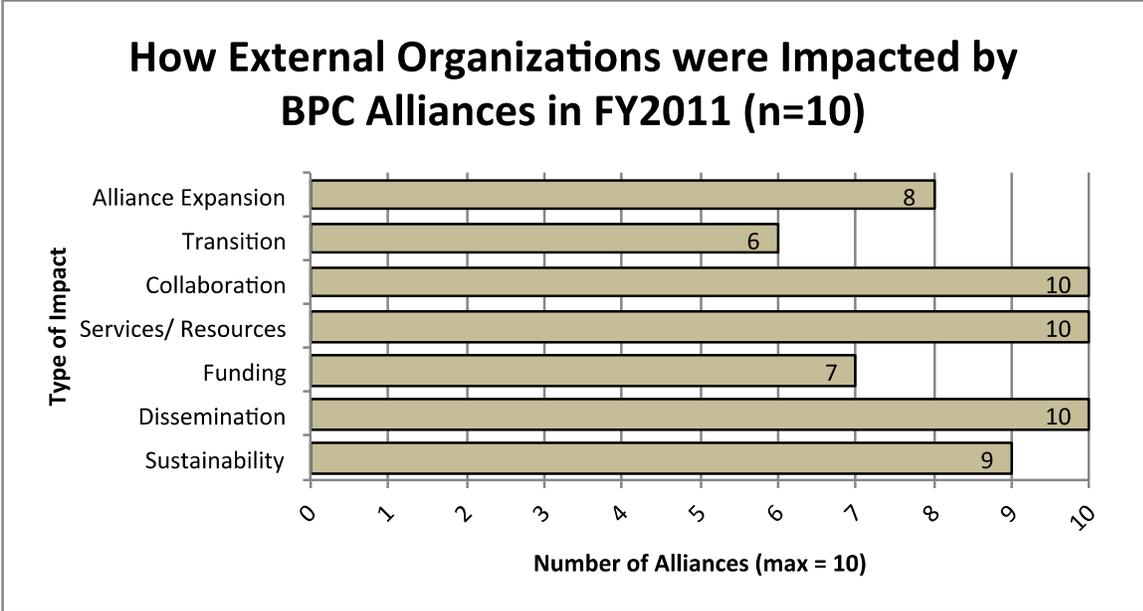


Figure 7

Finally, the BPC evaluators sought to measure the development of community and networks within projects supported by the program. This proved exceedingly difficult to separate from measures of either the individual or organization. However, the program did collect data on the extent to which projects collaborated with each other, an intentional aspect of the program. Figure 8 is a sociogram showing the intentional and reciprocal relationships built across projects.

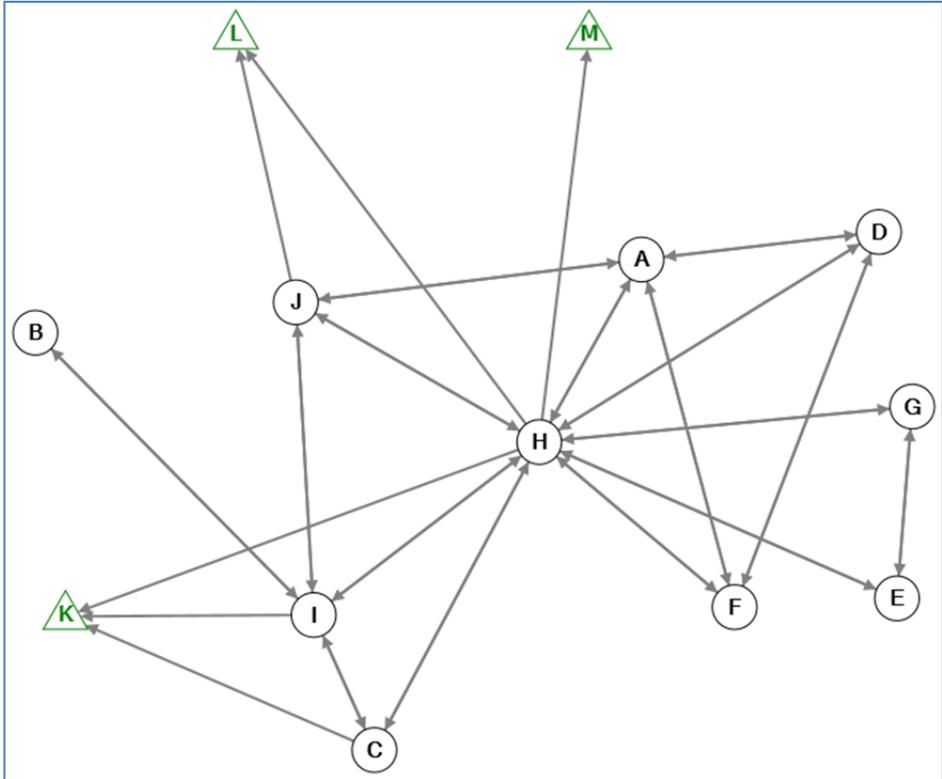


Figure 8

Discussion

Consolidating and Collapsing Data

Perhaps one of the most difficult aspects of the exercise in describing a body of programs is collapsing data across all programs into a meaningful report. This requires collapsing descriptive data of basic program characteristics (like the number of participants), collapsing outcomes, and collapsing the link between program activities or characteristics and outcomes data. The exercise in collapsing data looks a bit like qualitative research. It requires learning about each individual project (almost ethnographically) within a program, deriving an initial set of indicators with agreement and input from the project evaluators and PIs, analyzing early data from those indicators to see exactly what data come forth, re-defining the indicators to more closely match the actual data while maintaining data integrity, presenting findings to the group of evaluators and PIs, and showing the gap, if any, between the data collected and the extent to which the broader evaluation questions have been answered.

Evaluation Focus

While it is critical to involve project evaluators and PIs in the development, analysis and reporting of common, core indicators, the project PIs and evaluators are primarily focused on evaluating their own projects. There is benefit to inviting an external organization to coordinate the common, core effort in that they can devote their energy to the primary purpose of answering evaluation questions central to the program. The external organization can also work to clarify the critical evaluation questions and translate those to the project evaluators. Theoretically, all involved want to align their efforts to answering the critical questions of the program officers which most likely also answer the most critical questions of the individual projects. External, program-level evaluators should have the time and resources to focus on a central set of the most critical program-level evaluation questions, communicate that focus to the project evaluators, solicit their support, and begin aggregating and consolidating indicator data. They have the ability to organize and coordinate that individual project evaluators may not be able to provide. Plus, project evaluators are primarily focused on their own, specific project evaluations and only secondarily focused on the larger program evaluation.

Challenges

Challenges abound in any evaluation, and they are compounded when trying to evaluate multiple projects to describe the merit and worth, or even justify the existence of, a program. Certainly, one challenge is defining indicators for which data across projects may be collected. It is easy to simply count participants and stop there, since very little beyond participation is common across programs. Second, one of the most challenging aspects of this work is the need to track students through transitions. Many state data systems do not track students from high school into post-secondary institutions. Many evaluators have tried using social media to keep track of students after completing a program. These efforts are promising and certainly better than maintaining a database of participant

contact information and following up on a regular basis. This is notoriously difficult and requires the best thinking among the evaluation community to reliably and efficiently track participants. Third, evaluators are rarely a part of the reporting conversation that happens between program officers and the decision-makers to whom they report. Seeing first-hand how the data are used would aid the evaluation community in building more effective indicators and reporting mechanisms. It also helps us to see that the request is valid instead of a futile and expensive exercise in report generation.

Evaluation Use

While our ability to collect, analyze, and report data may increase as stakeholders make greater demands for rigorous evaluation, it does not follow that agencies, stakeholders, and other decision-makers will make greater use of these data. In fact, Patton (1997) describes a long history of inaction in the face of evaluation findings imploring action. What activities must evaluators take to increase the likelihood that decision-makers use program-level findings?

Limitations

First, this paper extracts only the indicators from the framework documents of both Broadening Participation and Informal Science. The documents are much more comprehensive and well worth reading. Also, other sets of common, core indicators exist, and this is not intended to be a meta review of indicators. Instead, this describes two very different sets of indicators both emanating from the same agency. Second, the figures from the Broadening Participation in Computing analysis are not presented as best practices. Instead, they exemplify how one group of evaluators tackled the problem of data consolidation and stand as an invitation for improvement. Finally, the author of this paper is naturally biased toward developing indicators as part of a team of project evaluators since this has been his predominant experience.

References:

- Clewell, B. & Fortenberry, N. (Eds.). (Jun 30, 2009). Framework for Evaluating Impacts of Broadening Participation Projects. (Available at: http://www.nsf.gov/od/broadeningparticipation/framework_evaluating_impacts.jsp)
- Friedman, A. (Ed.). (March 12, 2008). Framework for Evaluating Impacts of Informal Science Education Projects [On-line]. (Available at: http://insci.org/resources/Eval_Framework.pdf)
- Jones, R. (2012, May 23). House rejects move to cut \$1.2 billion from FY 2013 NSF appropriation. American Institute of Physics. Retrieved October 12, 2012, from <http://www.aip.org/fyi/2012/071.html>.
- Mervis, J. (2011, September 14). Senate Panel Cuts NSF Budget by \$162 Million. Science Insider. Retrieved October 15, 2012, from <http://news.sciencemag.org>.
- Patton, M. (1997). Utilization-focused evaluation : the next century text. (3rd ed.). Thousand Oaks Ca: Sage.

- Presidential Debate 2012 (2012, October 3). [Television Broadcast]. Denver, CO: CBS News.
- U.S. Department of Education. (2007). Report of the Academic Competitiveness Council. <http://www.ed.gov/about/inits/ed/competitiveness/acc-mathscience/index.html> (accessed December 22, 2011).
- Vastag, B. (2011, November 21). NASA and National Science Foundation are spared from big budget cuts - The Washington Post. Retrieved February 26, 2012, from http://www.washingtonpost.com/national/health-science/nasa-and-national-science-foundation-are-spared-from-big-budget-cuts/2011/11/16/gIQAz7FwhN_story.html
- Wenger, E., B. Trayner, and M. de Laat. 2011. *Promoting and Assessing Value Creation in Communities and Networks: A Conceptual Framework*, Rapport 18, Open Universiteit, Ruud de Moor Centrum.