

---

**OBSERVATIONS ON PFI EVALUATION CRITERIA**

---

In light of the NSF's commitment to measuring performance and results, there was strong support for undertaking a proper evaluation of the PFI program. The preceding chapters reported on a number of specific outcome measures and process measures (signposts) that were suggested to assess program performance; we now turn to some of the broader principles that were articulated (or could be inferred) from the workshop discussions.

**SOME GENERAL PRINCIPLES FOR EVALUATION**

Underlying beliefs about the fundamental nature of the innovation process and the characteristics of successful partnerships importantly shaped views regarding appropriate selection and assessment criteria for innovative partnerships and the sorts of foundations that are necessary to ensure their sustainability. Participants posited a number of general principles that might guide the development of evaluation criteria for the PFI program:

- It is important to measure both innovation and partnerships.
- There are important differences between *program* evaluation criteria, i.e., criteria by which the benefits and costs of the overall PFI program should be judged, and *project* evaluation criteria that are needed to generate data for the overall program assessment. Program evaluation criteria should be established before project evaluation criteria and should explicitly tie project-level measures to the larger program evaluation.

- There are important differences between *outcome* measures that would be measurable only at the end of a partnership and *process* measures that could be used while the project was under way.<sup>1</sup> Outcome measures should be established before process measures.
- In all cases, participants favored the development of “objective” measures of evaluation, recognizing the difficulties in developing such measures.
- Benchmarking and establishing indicators as a standard of comparison need to be done at the beginning of a project to establish a continuous data flow and ensure that relevant data are not lost.
- Finally, workshop discussants asked who (e.g., the PIs themselves or independent evaluators) should perform the outcome evaluations and when? Although it is important that PIs provide input on the performance measures that will be used, most supported independent, paid evaluators.

## OUTCOME MEASURES

There was some agreement on the principles that should guide the development of metrics, even if there remained a number of unresolved questions:

- Many favored development of a “strong logic” for the choice of goals and metrics to assess their accomplishment, while providing the flexibility necessary to enable application to a wide range of projects.<sup>2</sup>
- The PFI program’s multiple program goals has led to diversity in the PFI portfolio of projects, which also promotes multiple project goals and makes it difficult to make comparisons. Although discussants agreed that not every project needed to fulfill each

---

<sup>1</sup>For example, outcome measures were viewed as end results and summative in nature, useful for assessing goal accomplishment at the end of a partnership, whereas process measures were seen as real-time and formative, useful for providing feedback for improving the program.

<sup>2</sup>The metrics used should depend on the goals and the emphasis of work (e.g., workforce vs. technical studies).

PFI goal, and that projects should be evaluated on the basis of what they originally proposed to do, questions remained as to whether multiple goals should be equally weighted in an evaluation.

- Benchmarks need to be established, even if those benchmarks are somewhat imperfect. Initially, one may be able only to establish benchmarks derived from the goals of the partnership and then to ask whether (or the degree to which) it achieved specific goals; these may change over time, but this provides an initial basis for comparison, or straw man.<sup>3</sup>
- Some respondents argued that it was more important to get answers to fundamental questions rather than more esoteric ones that might need to rely on sophisticated measurement techniques. For example, how was the partnership formed? How was the venture expanded? What were the specific goals to be achieved? Were the goals achieved?

In navigating these various imperatives and challenges, workshop discussions suggested a combination of top-down and bottom-up approaches that would need to be connected to perform both program- and project-level outcome assessments:

- On the one hand, participants supported a top-down approach in which program evaluation criteria first were established based upon program goals, and project-level outcomes mapped to these larger program goals;
- On the other hand, it was clear that judgments about the outcome of the overall PFI program rested upon a summation of judgments about the individual projects, and the heterogeneity of the projects suggested that a fair amount of tailoring might be required.

To illuminate some of these interdependencies, we now summarize views regarding the top-down program-level outcome measures that

---

<sup>3</sup>For example, a partnership might have the goal of funding 12 graduate students, with seven of them going to work in this area after graduation and the project self-sustaining in three years. At the end of the evaluation period, the partnership might have fallen short of its goals, but the initial goals would provide a seemingly reasonable initial point of comparison.

derive from PFI program goals and the sorts of summative judgments that will need to be made. We then discuss workshop participants' views on project-level outcome measures and how these project-level measures might be summed in such a way that they can be re-connected to the program-level evaluation.

### **Program-Level Outcome Measures**

Workshop participants argued that the PFI program should largely be evaluated on the basis of whether it achieved its goals, including stimulating the transformation of knowledge, sustaining innovation, and transferring technology; training and workforce development; catalyzing economic development of states and regions; and broadening participation in the innovation enterprise.

Leaving aside the inherent difficulties of developing outcome measures for these goals,<sup>4</sup> there also was some support for evaluating the program on the basis of several less easily measured outcomes. Among these were estimating the value added by the partnership, including the types of benefits, number of beneficiaries, and distribution of these benefits; the spread of ideas on how to partner effectively and ensure best practices; the sustainability and long-term effect of the partnerships; the amount of new knowledge gained; and whether the award increased the propensity of awardees to partner more in the future.

There also was some support for the notion that the outcome measures should be capable of informing decisions about how to improve the PFI program: to learn which things are working, identify bottlenecks and “failure nuggets” and remove them, and how to better perform risk analyses. There also was a strong sentiment that such evaluative information could facilitate the identification of what the most effective strategies might be and what might constitute best practices at the national level.

---

<sup>4</sup>One challenge in evaluating outcomes is the inherent difficulty of counterfactual analyses. For the PFI program, one needs to establish that the partnership would not have occurred and the stream of benefits that resulted from the partnership would not have been seen had the PFI grant not been provided. To do this, one would have to interview the proposers of those projects that were not funded and discover whether those partnerships were in fact established even without PFI funding.

## **Project-Level Outcome Measures**

Although workshop participants felt that project evaluation metrics should be tied to the PFI program's stated goals, they also recognized that the partnerships were of many different types, promoting different mixes of the program goals. The key issues regarding outcome evaluations at the project level, then, were determining which goals applied to which project and what combination of metrics accordingly should be used to evaluate goal accomplishment for each, so that these results could provide the data necessary for the program assessment.

Additionally, some expressed the view that at the project level, projects could (and should) be evaluated largely on the basis of whether they had met the specific goals detailed in the original proposal to the PFI. To this end, it was believed that principal investigators should initiate conversations with the NSF over which criteria would be used to assess which outcomes of their partnership.<sup>5</sup>

## **PROCESS MEASURES**

### **Program-Level Signposts**

The program-level signposts were seen as summative, based upon project-level measures, described next.

### **Project-Level Signposts**

As was described in preceding appendixes, attendees advocated project-level process measures (signposts) of two general kinds. The first kind focused on the progress of the partnership in achieving its innovation goals (which could indicate whether partnership activities were leading toward results); the second focused on the characteristics of the partnership that could indicate whether the partnership was using available instruments in pursuit of the correct actions.

---

<sup>5</sup>Although the idea may have some merit, there was no suggestion that PFI applicants should include in their application a suggested set of evaluation measures for their project.

## EVALUATING THE NSF'S OWN SUPPORT

Finally, consistent with the view that the evaluation process presented the NSF with an opportunity to turn the lens on itself in an effort to improve its support for innovation, attendees suggested that the partnerships should be asked to evaluate the NSF's support during the PFI program and the extent to which that support helped their partnership. It was unclear how best to acquire this information (e.g., by open-ended responses or a by standardized rating scheme), but the basic aim was to elicit partners' views on such questions as whether the program had been a success and worthwhile experiment from their vantage point, whether they would recommend the program to others, and what they learned about bootstrapping the innovation process. The project-level evaluations of the value of the NSF's support also were viewed as summative and easily aggregated for the program as a whole.

Some also saw program-level process measures as the place for the NSF to turn the lens on itself regarding selection and grant-making criteria, the quality of the technical assistance it provides, and other issues.<sup>6</sup> The quality of the technical assistance provided by the NSF can be summed from the individual project-level assessments of how appropriate and helpful that technical assistance was.

Workshop participants identified a wide range of challenges related to the methodology for evaluating individual projects and the program as a whole and in some cases offered practical suggestions regarding how best to implement an evaluation of the PFI program. We now consolidate and summarize these observations and suggestions.

- *The Difficulties of Evaluation in Observational Studies.* One challenge arises from the fact that a PFI evaluation would not be a controlled experiment<sup>7</sup> but an observational study, requiring

---

<sup>6</sup>For example, the time it takes to make an award was seen as important.

<sup>7</sup>In other words, projects were not randomly selected to receive a treatment (funding), which would have enabled a comparison between those that were funded and those that were not.

careful handling of potential bias.<sup>8</sup> To evaluate outcomes in such a setting, one needs to compare the outcomes of the chosen partnerships with some referent, but it was somewhat unclear exactly what that referent should be. Should a partnership's experience be compared to a counterfactual where the partnership did not receive PFI funding? What sort of confidence could one have in such a comparison? If partnerships are to be compared to projects whose proposals were not selected by the PFI program, how should the inherent biases be handled?

- *Self-Evaluation Versus Independent Evaluators.* There was support for including stakeholders in determining how their projects should be measured, but there was somewhat mixed support for the broader principle of self-evaluations. On the one hand, some considered the prospect of self-evaluation as presenting an opportunity that might be abused; others perceived a financial involvement that lent credence to self-evaluation and straightforward numbers that partners can provide. In light of the concern that self-reported measures could be biased, there was broad support for bringing in an external evaluator; the point also was made that NSF programs require a paid, outside evaluator for all projects.
- *Maintaining Records.* Another challenge was collecting and maintaining the necessary records to inform the evaluation; it is important that partners establish a process early in the project for collecting and maintaining records that can provide the NSF with relevant performance information. Some argued that incentives might be established, for example, tying the evaluation to the collection of information that also could be used later in marketing the project. For others, the main purpose was not so much to evaluate a particular project but to evaluate the overall program. This information was seen to be useful in coming to conclusions regarding what works and what does not and for supporting recommendations to policymakers so that program design can be improved.

---

<sup>8</sup>The selected partnerships were those that appeared to have the most promising proposals, and those that were not selected had proposals that were, in some sense, judged inferior.

- *Time Horizon for Evaluation.* Some thought that the time horizon of the PFI program was too short to be able to ascertain the program's effect on innovation and sustainability. For them, it was just too short a time scale for an evaluation. Some suggested a "3+2" year evaluation period—three years for the initial support, with two years of follow-up on outcomes. By the same token, a two-to-three-year time horizon was seen as probably being long enough to evaluate the NSF's role as a catalyst.
- *Data Reliability.* Several workshop participants expressed concern about the comprehensiveness, validity, and reliability of performance data.<sup>9</sup> Observing that data can be no more accurate than the way it is reported, it is important to make data collection as simple as possible. The observation was made that too many studies hide behind a great deal of complexity, and it is quite difficult to establish that the underlying data are even valid. Given that the PFI program is small, it should be relatively easy to interview all participants.

---

<sup>9</sup>They cited the case of the SBIR reauthorization as evidence of the importance of reliable and valid data.