15

Establishing System-Wide Evaluations: Program-Centered Evaluation and The National Science Foundation's Industry-University Cooperative Research Centers

Denis O. Gray

Denis Gray is an Associate Professor in the Department of Psychology at North Carolina State University in Raleigh, North Carolina.

Abstract

The new generation of "partnership-based" technology initiatives, including the Manufacturing Extension Partnership program, invariably involve a variety of stakeholder groups, including: federal and state-level sponsors, local institutions and their program staff, and customers. Since each group tends to have different information needs, the ideal evaluation should incorporate diverse evaluation approaches including, monitoring, summative, and formative evaluations. Unfortunately, a variety of forces tend to produce evaluations which are "sponsor-centered", that is, they emphasize the information needs of sponsors, often to the neglect of program staff and customer needs. This paper will describe a technology program evaluation which has avoided this problem by stressing a "program-centered" approach.

The National Science Foundation's (NSF) Industry-University Cooperative Research Centers (IUCRC) Program attempts to promote research development and technology transfer by creating university-based industrial consortia. The IUCRC's program-centered evaluation emphasizes: flexibility, a focus on critical program processes and outcomes and timely feedback of information. The overriding goal of this element of the evaluation effort is to aide continuous improvement.

The linchpin of the IUCRC evaluation system is the on-site, local evaluator. This individual is responsible for collecting qualitative data via observation and quantitative data via a standardized process/outcome questionnaire. The "real-time reconnaissance" obtained through the participant-observer role can be feedback quickly and informally and serves as an "early warning system". Use of feedback based on the standardized process/outcome questionnaire has facilitated national benchmarking and identification of empirically-based "best-

practices" on a variety of dimensions including: satisfaction with research and administration and technical and commercialization outcomes.

The IUCRC evaluation effort demonstrates that program, as well as sponsor, information needs can be met within the same evaluation effort. For the IUCRC program the payoffs from this approach include the active support of program staff for the evaluation, a climate which encourages and reinforces continuous improvement, and feedback from "customers" which indicate the program continues to meet their needs. Efforts to support continuous improvement will be enhanced with the imminent release of an "management handbook", another spin-off of the evaluation effort.

Introduction

This paper will be different from many of the others which will be delivered as part of this workshop. Instead of focusing on the results of our evaluation efforts, I will focus on how and why a "program-centered" evaluation approach has been used to evaluate a multi-site federally-funded university-based technology development initiative. In doing so I will try to highlight the implications of this experience for manufacturing modernization programs evaluation policy and practice.

Evaluator's Reverie?

During the winter of 1989, something quite extraordinary, some might say unprecedented, in the annals of program evaluation happened during the annual directors meeting for the National Science Foundation's (NSF) Industry-University Cooperative Research Centers (IUCRC) Program. While NSF staff were explaining why and how they must begin to phase out all operational support for the oldest cohort of centers, a director raised his hand and asked if NSF would consider continuing to support their center's participation in the IUCRC's program-wide evaluation effort? When NSF indicated they would consider such a request, the directors went into a closed session to discuss the pros and cons of this proposal. While history hasn't recorded the discussion which took place during that session, the record shows that the directors unanimously passed a motion requesting NSF continue to fund the evaluation effort at all such centers.

While this story may sound like the product of an deluded evaluator's over-active imagination to some, if not most, over-scrutinized technology program managers, the events described above actually occurred. In fact, seven years later all of the centers that were affected by this decision are still involved in the evaluation effort through a similar mechanism. Why did the IUCRC directors come to the defense of, and to this day continue to cooperate in, a responsibility most managers resist and others loathe? The fundamental answer appears to be that evaluations, if they are designed and implemented appropriately, can provide managers of technology programs with valuable information -- information that can be used to improve program operations and, ultimately, performance. And when evaluations do this, managers can become more than reluctant and passive participants in the evaluation process, they can become supporters and sometimes advocates of evaluation.

What kind of evaluation strategy is likely to produce such a result? Does it have any relevance to managers involved in manufacturing modernization programs (MMP)? In order to answer these questions, I will describe the IUCRC program and its "program-centered" evaluation system. In doing so, I will try to highlight the principles and procedures which have contributed to the success of this evaluation effort and which may generalize to the situations faced by MMP managers.

IUCRC Program and Model

The NSF IUCRC program began as part of the Experimental R&D Incentives program funded in the early 1970s and was formalized as an NSF program about 1979 (Burger, Cole & Burnett, 1982). It is currently administered by the Engineering Education and Centers Division of NSF's Engineering Directorate. The program has multiple goals (e.g., promote closer communication and collaboration between industry and university), however, it's most important goal is to develop and transfer new knowledge and technology to industry. NSF attempts to achieve this end by providing cost-sharing and technical assistance for the establishment of IUCRCs on college campuses.

A detailed description of the IUCRC model can be found elsewhere (Gray, Hetzner, Eveland & Gidley, 1986). In brief, an IUCRC is a university-based, typically. multidisciplinary, industrial research consortia. Since centers are housed at one or several universities, most research is performed by faculty and graduate students. IUCRCs tend to follow a relatively standardized set of policies and procedures, including: members pay an annual fee (usually between $30 to 50K/year) to support the center's research program; members get equal access to and ownership of all research and intellectual property; findings, know-how and technology are transmitted through a variety of means including periodic reports and semi-annual meetings; and members get one vote on the center's Industrial Advisory Board (IAB). However, in contrast to most industrial affiliate programs, membership has its privileges: IAB members exercise *de facto* control over planning and selection of the center's program. As a consequence, most of the center's research can probably best be described as strategic or preproprietary fundamental research. This is, the research, while fundamental, has been selected because of its direct relevance to industry's needs. It's worth noting that the IUCRC program was the model upon which a number of state "center of excellence" programs were based. Thus, it should be viewed as a technology development program, rather than a technology deployment program.

A review of the findings on the IUCRC program are beyond the scope of this paper. But it's worth noting that the program appears to be successful by a number of criteria: (a) compared to other technology development programs the IUCRC has done more with less. Specifically, it only receives about US$4 million/year from NSF, but boasts over 50 centers, involving 75 universities, with 750 researchers, 1000 students, is supported by over 600 firms, with a total budget of over $66 million. (b) Member and faculty satisfaction are high. (c) Firms report a high level of knowledge and technology transfer. In the view of some observers, the program's evaluation effort has contributed in at least some small way to this success.

Relevance for Manufacturing Modernization Programs

While the goals (e.g., technology development vs. technology deployment), customers (e.g., large R&D intensive firms vs. small often low tech firms) and operations (research vs. a variety of service, consulting and training activities) of an IUCRC and most MMPs are quite different, the programs also have a number of things in common. Over half the IUCRCs conduct research and transfer technology which is directly or indirectly related to the manufacturing process, many involving applications to traditional manufacturing areas. For instance, the Georgia Institute of Technology center conducts work on materials handling and the University of Connecticut center performs research on grinding and related technologies. In addition, many IUCRCs receive substantial support from state-level Science and Technology (S&T) or economic development programs and

are held accountable for the same kinds of outcomes MMPs are (e.g., job creation and retention; local economic impact). More importantly, both types of programs face a similar set of circumstances which affect their evaluation strategies. These circumstances represent both challenges and opportunities for conducting sound evaluations and include:

- Evaluation must take into account the needs and expectations of different stakeholder groups (e.g., federal and state government, industrial participants, institutional interests within their host organization);
- Stakeholder groups often have different value systems, needs and time horizons when they approach the evaluation process;
- Desired outcomes are multiple, temporally and physically removed and often proprietary, and therefore difficult to pin down;
- Evaluation is expected to occur in the context of a large multi-site national program;
- Desired outcomes are embedded in the complexities and uncertainties of the innovation process;
- The most important stakeholder, the "customer", operate in fast changing, highly competitive environments, where one's ability to master change often is a matter of survival;
- Desired effects are highly dependent upon the ability of management to orchestrate the operations of a complex multi-faceted organization;

Thus, since MMPs face similar circumstances, the IUCRC evaluation effort may have relevance to meeting their needs.

Balancing Program and Sponsor Evaluation Needs

In principle, evaluations can and should be designed to meet multiple needs of various stakeholders including program sponsor(s), program manager and program "customer" (Rossi & Freeman, 1993). Since sponsors are typically interested in accountability and program justification issues, their evaluation needs are usually best met through monitoring and summative evaluation methods. *Monitoring evaluations* usually address, so-called bean-counting, questions like: Did you do what you said you were going to do? Did you serve who you said you would serve? *Summative evaluations* usually address questions like: Did the program achieve it's ultimate objective for the *whole* target population. Since managers and customers are interested in whether operations are working as they should *right now*, their evaluation needs are usually best met through formative or process evaluations. *Formative or process evaluations* typically address questions like: Does this procedure achieve the intended immediate outcome? How can we improve this procedure so we obtain the intended short-term outcome?

Without any doubt the ideal evaluation of a technology initiative will incorporate all of these approaches (Solomon & Tornatzky, 1986). Unfortunately, a variety of forces tend to undermine the development and implementation of balanced multi-stakeholder (and therefore multi-method) evaluations. First, evaluation resources (both money and time) are limited, so it sometimes seems difficult to adequately address all needs. Second, evaluations are paid for and tend to be driven by what the sponsor thinks is important. As a consequence, most evaluations tend to be sponsor-centered. That is, they are driven by sponsor information needs. Unfortunately, at the same time, the information and evaluation needs of local program managers

and customers tend to be neglected or at best become an after-thought. These problems tend to be magnified in high-profile, multi-sponsor and controversial programs like the new generation of innovation/modernization programs.

There are a variety of undesirable consequences to relying on such a system: program managers perceive and respond to evaluation requirements as both an intrusion and a burden; collected information is often irrelevant or too late to meet local needs; information which could affect local program performance is not collected. In short, as a recent survey of industry participants in a state/federal technology development program confirms, such evaluations do little to meet the needs of program managers and offer little value to customers. Since we're unlikely, and ill-advised, to disregard sponsor needs, the antidote to this situation is to balance your evaluation effort by insuring that your evaluation includes a program-centered evaluation approach. In the next section, we describe an evaluation which has strived to create such a system -- the IUCRC evaluation.

Designing a Program-Centered Evaluation

A detailed description of the entire evaluation process can be found elsewhere (Hetzner, Gidley & Gray, 1989; National Science Foundation's Industry-University Cooperative Research Centers (IUCRC) Evaluation Project, 1996). While the evaluation protocol has evolved over time, it has been consistent in attempting to address sponsor, university/faculty, manager and customer evaluation needs through a combination of monitoring, summative and formative evaluation assessments. The evaluation system includes two major components: a structured ongoing evaluation and a collection of supplemental focused studies. A graphic representation of the assessment process is presented in Figure 1. Since the ongoing systematic evaluation has been designed to be "program-centered", we will focus on this component.

What is program-centered evaluation? A program-centered evaluation attempts to provide information which can be used to refine and improve program performance. In a sense, it is a combination of what the literature refers to as formative and process evaluation (Rossi & Freeman, 1993).

Like all good evaluations, program-centered evaluations need to embody a number of basic qualities, including: objective, systematic, and use of diverse and complementary methodologies. However, since program-centered evaluations attempt to aide program performance, they must possess some other qualities, including: flexibility, a focus on critical program processes and outcomes (typically nearer-term), timely feedback, and an emphasis on continuous improvement. In essence, program-centered evaluation shares a lot of things in common with practices associated with total quality management. A major challenge then becomes, how to engineer a program-wide multi-site evaluation which also meets decentralized needs for locally relevant information.

In the next, section I will describe the structures and processes of the IUCRC evaluation which embody these principles.

IUCRC Evaluation System

The following features define the structural elements of the multi-site evaluation system.

On-site evaluator

The linchpin of the IUCRC evaluation is the on-site evaluator. This individual is responsible for executing the NSF-sponsored evaluation protocol at each site. Major

qualifications for this individual include a background in evaluation research (or some other area of applied social science) and no direct association with the academic unit which operates the center. The first qualification insures competence and professionalism and the second a degree of objectivity.

Most evaluators, particularly in summative studies, are expected to be nothing more than objective and disinterested observers. However, in order to be effective, the program-centered evaluator must stray from this role Rossi and Freeman (Rossi & Freeman, 1993) describe the required modification in this way: "Evaluators engaged in formative studies obviously must become involved in the actual design and programming effort, since the emphasis here is on increasing the success of subsequent intervention efforts and their evaluations. Thus, the evaluator becomes an advocate and a partisan participant in the program activities" (p. 137). Consistent with their recommendations, IUCRC evaluators are expected to serve as organizational development consultants, "fair witnesses" and, yes, sometimes advocates.

Since the evaluator must strive to be *both* objective and involved, there are plenty of opportunities for conflict and ambiguity in fulfilling this role. In this vein, I find it interesting and probably appropriate that on organizational charts NSF usually shows the evaluator linked to the center with a solid line and to them with a dotted line, while centers show just the opposite!

Assessment/Instrumentation

In order to meet the needs of the program, the assessment tools used as part of the systematic ongoing IUCRC evaluation system have evolved over the past fifteen years. For instance, a number of tools are no longer used (e.g., network analysis, student assessment), and most others have been revised several times. Nonetheless, the assessment still involves collection of both qualitative and quantitative information. The protocol currently used includes three formal instruments: historical profile, the process/outcome questionnaire and the exit interview (National Science Foundation's Industry-University Cooperative Research Centers (IUCRC) Evaluation Project, 1996).

Historical Profile. Evaluators are expected to be participant-observers of significant center activities and transactions and to summarize their observations and conclusions in an annual historical profile, in essence a qualitative case study.

Because of budgetary limitations participant observation is usually restricted to semi-annual research review/board meetings, internal management meetings, and access to all pertinent program documents (e.g., correspondences, progress reports, etc.). Since a great deal of the center's transactions take place during these events, this has proven satisfactory. The historical profile instrument attempts to focus the evaluator's attention on key aspects of center operations (e.g., changes in the environment; turnover of key staff; decision-making, etc.). Evaluators currently use an abbreviated version of much longer instrument which was originally developed for a multiple case-study analysis.

This element of the IUCRC evaluation is critical to the program-centered evaluation approach. First, it allows the evaluator to understand the center and the challenges it faces in a more holistic way. Second, it provides a basis for obtaining real-time reconnaissance on center operations and performance. And perhaps most importantly, it provides entree and credibility when the evaluator serves in his/her consultant/advocate role.

Process/Outcome (P/O) Questionnaire. A different version of this instrument is administered annually to two important stakeholder groups, industry and faculty. These instruments have undergone extensive revision over the past ten years based on feedback from directors and the experience of a number of

evaluators and provide both quantitative and qualitative data on performance. Constructs addressed in the current questionnaire were selected based certain criteria: they represent important program processes which are believed to influence short and intermediate outcomes; they represent short and intermediate outcomes which are believed to influence longer-term (summative) outcomes.

There are three main components of the questionnaire: process items, outcome items and open-ended items. Industry process items use a forced-choice likert-style satisfaction format (1 = not satisfied; 5 = very satisfied) to assess industrial member's satisfaction with both the center's research program (e.g., quality, relevance, focus) and administrative operations (e.g., communication, project selection, technology transfer). Industry outcome items use a forced-choice likert-style format (1 = no impact; 5 = very high impact) to assess the impact the center has had on R&D (e.g., accelerated or improved existing research projects; dollars invested in follow-on projects), commercialization (e.g., development of new products, processes or services) and non-technical (e.g., student recruitment) outcomes. Global satisfaction and an intention to renew membership round out the instrument. Because enhanced performance at the local level is central to a program-centered evaluation, each process and outcome section is complemented with an open-ended question which solicit feedback on how the center can improve satisfaction and/or impact. This mechanism has worked very well. During our last assessment, 300 respondents provided more than 600 codeable comments.

The P/O questionnaire insures that the evaluation is "customer-driven" and "continuous-improvement" oriented. Annual collection of data provides a basis for measuring change in performance over time. If the data collected with this device is used correctly (see "Evaluation Process" below), it reduces the need for the last assessment tool.

Exit Interview. This brief telephone interview is used with organizations who terminate their memberships. It is designed to provide information on why members left the center and what changes the center management could or should make to prevent or undo their resignation.

Coordination/Aggregation/Bench-marking

The IUCRC evaluation system is intended to serve both the oversight needs of the sponsor, NSF, and the local information needs of over fifty centers. These tasks require coordination and support. These needs are met through the efforts of a team at North Carolina State University (NCSU).

In addition to collecting structural data (e.g., funding, membership, staffing) directly from centers and coordinating the activities of local evaluators, the NCSU team provides aggregate analyses for NSF and support the local evaluators feedback function. In the context of a program-centered evaluation, a critical aspect of the latter function involves providing "normative data" for the P/O questionnaire. This allows local evaluators to provide "benchmarking" feedback to their center directors.

IUCRC Evaluation Process: Emphasis on Feedback and Improvement

As I've stressed throughout this paper the key to program-centered evaluation is providing local management with information which can be used to refine and improve program performance. This objective is met through informal and formal feedback to center management.

Informal Feedback. Managers, particularly in programs affected by the quick pace of technological change, competitive pressures and the whims of congressionally and legislatively-mediated funding agencies, need to

make to adjustments quickly in order to survive. Unfortunately, structured formal evaluations are not known for their quick turnarounds. Many a program was going out-of-business or was out-of-business by the time the evaluator report was delivered saying they were in trouble! As a consequence, the IUCRC evaluation has tried to stress the importance the "real-time reconnaissance" obtained through the participant-observer role and informal feedback provided by the evaluator in his/her role as "organizational development consultant".

While we have not documented this activity carefully, directors have provided countless anecdotes about instances where evaluators provided critical information about: members who confided they were unhappy about some aspect of center management or felt a technical or other needs weren't being met or were on the verge of terminating their membership; faculty members who felt they were being asked to sacrifice quality for expedience. By receiving feedback quickly and informally, directors have an opportunity to correct the problems before serious harm is done (e.g. loss of a member). In other instances, evaluators have served as the channel for communicating "best practices" from other centers. I am convinced this aspect of the evaluation has proven helpful to local centers.

Survey-feedback. While formal structured feedback can be a bit slow, it can provide information that would not be available through other means. In the case of the IUCRC evaluation, the P/O assessment is really designed as part of "survey-feedback methodology" that is frequently used in industry (Dunham & Smith, 1979), and lends itself readily to the multi-organizational comparisons (Johnson, 1996).
The IUCRC Survey Feedback Cycle is graphically depicted in Figure 2 and involves four steps. Step 1: P/O questionnaire data are collected from industry and faculty; Step 2: NCSU aggregates data and provides "normative" and other statistics for benchmarking; Step 3: Local evaluators provide feedback to their centers, benchmarking them against last year's performance and IUCRC program benchmarks. Step 4: Center management can use the benchmarking information and the comments provided by their membership to refine and improve program operations and, hopefully, performance.

By using the data in this way, a standardized multi-site evaluation can operate in a truly decentralized fashion and help meet local information needs. It's worth noting that the ability to benchmark oneself against other centers, also allows one to use this information to identify best-in-class centers and highlight "empirically" supported best-practices.

Supporting Continuous Improvement

In my experience, one of the weak links in most program-centered evaluations or continuous improvement approaches is the final step -- figuring out what you should do better or different. Until recently, the IUCRC Program has addressed this need informally through a variety of mechanisms. First, the NSF program manager, who has over twenty-years experience in building and supporting local centers, really serves in a technical assistant capacity. Second, directors meet once a year and discuss common problems and "best practices". Third, evaluators meet twice-a-year for the same purpose and to discuss research findings and trends. However, a new piece of the puzzle will be added later this year, when a handbook, Managing the cooperative research center, is published later this year. This volume was authored by a team of evaluators who have cataloged and document what has appeared to be IUCRC "best practices".

Discussion

The IUCRC Evaluation system has proven that it's possible to build and implement an evaluation approach which emphasizes the needs of local program managers -- a program-centered evaluation. The benefits of the evaluation are probably harder to quantify and attribute. One indicator may be that the IUCRC evaluation has been cited as a successful evaluation effort in the literature (Roesner, 1989). Another indicator might be the success of the program. The IUCRC Program, leverages NSF support almost 20-to-1, and has an enviable survival record among it's centers. In my opinion, the evaluation approach deserves at least a little credit for this success. In addition, and perhaps, more importantly, the evaluation continues to enjoy the support and cooperation of the local directors.

References

Burger, R. M., Cole, S. H., & Burnett Jr., C. H. (1982). Analysis of the NSF's University-Industry Cooperative Research Centers Experiment: An effort to stimulate university-industry cooperative research in solving R&D problems in American industry. In R. M. E. Colton (Ed.), *Analyses of five NSF experiments to stimulate increased technological innovation in the private sector.*, . Washington, D.C.: National Science Foundation.

Dunham, R. B., & Smith, F. J. (1979). *Organizational surveys: An internal assessment of organizational health*. Glenview, IL: Scott, Foresman and Company.

Gray, D. O., Hetzner, W., Eveland, J. D., & Gidley, T. (1986). NSF's Industry-University Cooperative Research Centers Program and the innovation process: Evaluation-based lessons. In Gray, D.O., T. Solomon, & W. E. Hetzner (Eds.), *Technological innovation: Strategies for a new partnership*, (pp. 175-193). Amsterdam: North Holland.

Hetzner, W. A., Gidley, T. R., & Gray, D. O. (1989). Cooperative research and rising expectations. *Technology in Society ,11*, 11.

Johnson, R. H. (1996). Life in a consortium: The Mayflower group. In I. E. Kraut (Ed.), *Organizational surveys: Tools for assessment and change*, (pp. (pp. 285-309)). San Francisco: CA: Jossey-Bass.

National Science Foundation's Industry-University Cooperative Research Centers (IUCRC) Evaluation Project. (1996). *Evaluator's Handbook: National Science Foundation's Industry-University Cooperative Research Centers program*. Raleigh, North Carolina: IUCRC Evaluation Project.

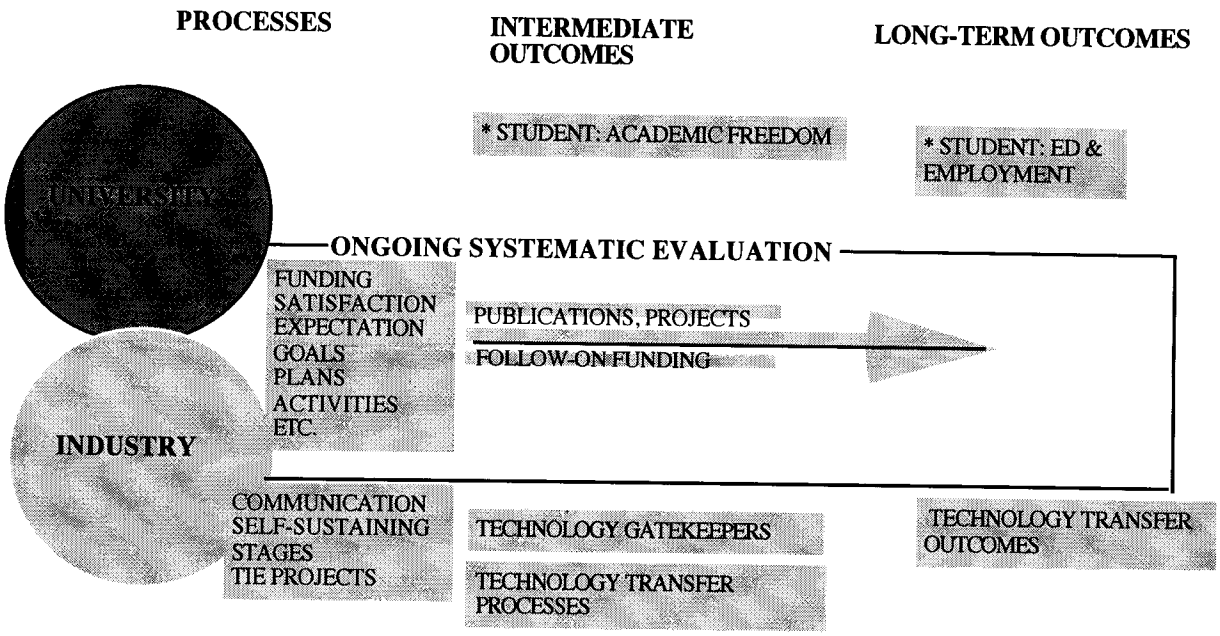
Roesner, J. D. (1989). Evaluating government innovation programs: Lessons from the U.S. experience. *Research Policy, 18*, 343-359.

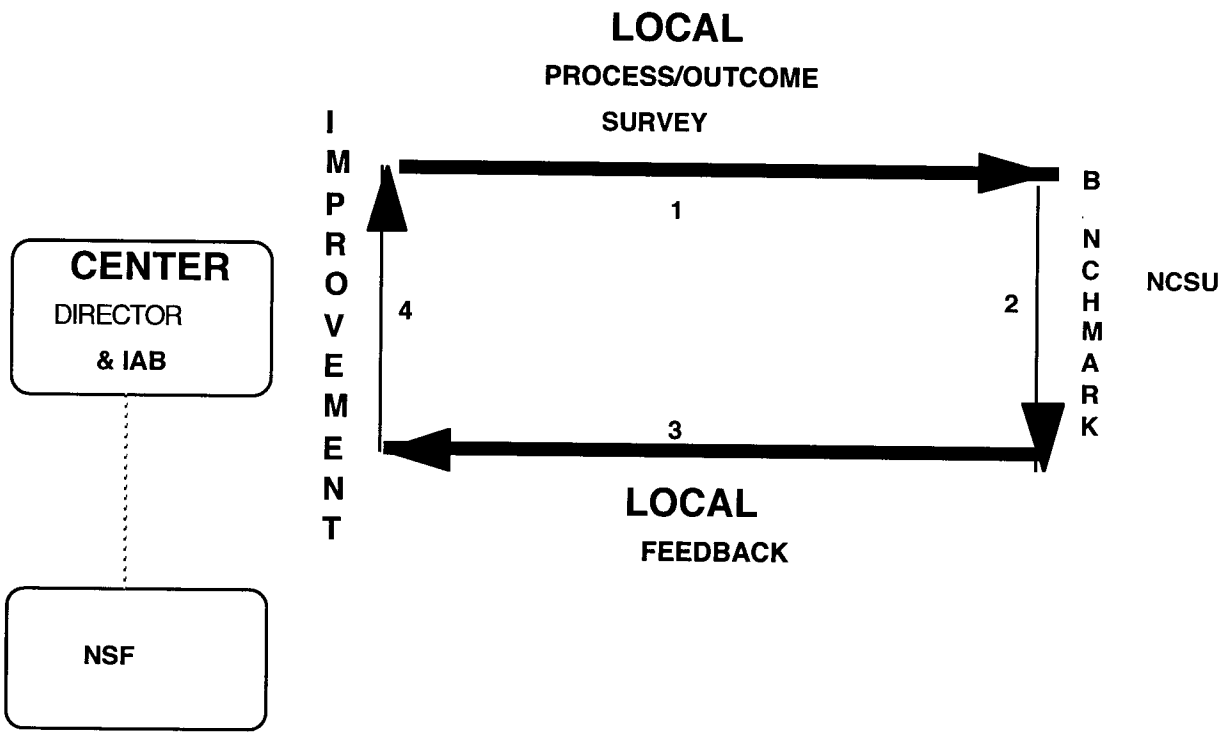
Rossi, P. H., & Freeman, H. E. (1993). *Evaluation: A systematic approach*. Newbury Park, CA: Sage.

Solomon, T. S., & Tornatzky, L. G. (1986). Rethinking the federal government's role in technological innovation. In D. O. Gray, T. Solomon, & W. E. Hetzner (Eds.), *Technological innovations: Strategies for a new partnership*, (pp. (pp. 41-53)). Amsterdam: North-Holland.

Figures 1 and 2

NSF IUCRCEVALUATION SYSTEM





Differing Values in Program Justification: Discussion

Oldsman. I have a comment about the results in the ROPI paper by David Burress. There were significant differences between the two panels' results. What does that tell you about whether or not one can reach consistent goals?

Sears. In the ROPI paper you talk about using the legislature to decide which measures to use. A major problem with this approach is that for many of the programs we are talking about the appropriate evaluation time frame is much longer than two years. But, because of the election cycle, the time horizon of most legislators is often only two years.

Burress. The weights could change drastically over time, but not the list of measures.

Luria. Despite the differences between the weights associated with the measures, it seems to me that both panels emphasized elements of value-added, such as income growth.

Feller. I've believed for many years in focusing on value-added. The problem for empirical studies is that most state governments are still looking at jobs.

Oldsman. It seems that this focus on jobs is amenable to manipulation, an educational process. The people in David Burress's panels were, over time, exposed to education. They arrived at a conclusion.

Shapira. One role of evaluation is that it is an educational process which helps clarify goals. MEP evaluations should have an educational element.

Oldsman. We have a conceit of knowledge, thinking that we are right. But other people come in with their own background and experience. A center director asked, "how do we define productivity?" Here is an important participant who doesn't know what productivity is. An explicit objective of evaluators is education.

Ellington. I am concerned that there is a disconnect with MEP center directors. In talking about how we structure evaluations, we must get the center directors' buy-in.

Burress. It's not our job to decide for them the goals of the program. The taxpayer gets to decide the goals of the program.

Ellington. My thinking has changed regarding who the audience is for this evaluation. There is a group of folks that think evaluation is a four-letter word. That sets up a potential problem. I believe we need to broaden the goals of the program beyond jobs and income. But politicians cannot understand complex systems, so I was pleased that David Burress went down to about three different indicators from 10. I question the value of having public citizens on the panel as opposed to professionals trying to move modernization further. It's an interesting dilemma to consider the role of citizens versus. professionals. One insightful comment was the basis of comparison--what could you do with the money which is being invested in manufacturing extension--you could either give it directly to the citizens or run this program. That's not a bad model. It's the way conservatives look at things.

The second paper, by Denis Gray, discusses the issues of multiple stakeholders and formative vs. summative evaluations. How much should be on monitoring vs. outcomes. It depends on the system's needs. We are a new program which might have more systems needs than say an SBDC or Agriculture Extension. The 1994 election shifted the emphasis to favor justifying the program, when under other circumstances we would have emphasized program improvement.

Rhoades. When I came here I expected more discussion about balancing economic impact and continuous improvement. I've been surprised at the lack of emphasis on continuous improvement. Too often the political side is banging on the door looking for program justification at a very early date and not giving enough time to work out operational issues.

Bury. We had a change in the MEP Evaluation Working Group. They said, "we want to change our focus to serving the needs of center directors." Now we are examining issues related to how available data can be transformed into something useful for the center directors.

Shapira. As we look at who is doing work, the weight has tended to be in program justification. However, there is some work on program improvement that didn't get aired at this workshop. Maybe the next stage of development for new MEP evaluation studies is to focus on continuous improvement. That said, we have made a lot of progress. I'm thinking back to the workshop in Atlanta in 1994 where John Redman spoke about the national system and introduced a series of new NIST evaluation initiatives. Many of the 1994 initiatives have produced the work we discussed today.

Haines. It doesn't bother me that we haven't spent time on program-centered evaluation. There is a lot of that via rollover and third year reviews designed with the idea of continuous improvement. To me the critical issue is to make that process better and link that to the program review process. We've now done 25 formal reviews. We'd like to try to mine some of those reports and do some cross-cutting analysis. We want to focus the review process and design to take into account what we are learning about the impacts of the results. We haven't set a standard of performance throughout the system. We should have a bar but we don't have it. We're collecting enough data to set the bar, or at least a range. I see us continuing to push the envelope in evaluation.

Regarding MEP program operations, we need to develop a generalized service delivery model for MEP so we don't have such a diffuse service model, a generalized model that the system is comfortable with. We may not specifically focus on distinctive firms, but we will have a set of "MEP" products. The reauthorization legislation will force us to discuss goals. How can we work together? Then we'll probably be changing metrics and decide what new metrics should be added, such as value-added. We've worked hard to get buy in and collect data from centers on the existing metrics. Once we make some more progress with collecting these existing metrics, we can have that discussion about what the new metrics should be.

I see us spending more time on communication with the centers. That will begin happening soon. If we run program reviews on results, centers will have to think about how they will better collect this information, which will lead to more emphasis on evaluation systems at the center level.

I don't know how we'll do this, but I understand that I will have to work with the states on evaluation. We have mechanisms, for example, our relationship with the National Governor's Association, but we haven't yet exploited those.

And lastly there's the little things, such as improving the case studies; and developing and implementing voluntary standards of best practice for evaluation among the centers.