

**EVALUATION OF  
THE  
APPALACHIAN  
REGIONAL  
COMMISSION  
INDUSTRIAL  
COMPETITIVE-  
NESS  
DEMONSTRATION PROJECTS**

**Melvin M. Mark,  
Irwin Feller, and  
Amy Glasmeier**

**Melvin Mark is Professor of Psychology at Penn State University, and a social psychologist and program evaluator. Irwin Feller is Director, Graduate School of Public Policy and Administration, and Professor of Economics at Penn State University. Amy Glasmeier is Associate Professor of Geography at Penn State University.**

This paper provides a selective overview of an evaluation plan for six industry information and worker training demonstration projects funded by the Appalachian Regional Commission (ARC) under its Regional Development Initiative to Enhance Manufacturing Competitiveness (Industrial Competitiveness Initiative). Some tentative generalizations are also made regarding the evaluation of manufacturing modernization more generally.

**The ARC's Industrial  
Competitiveness Initiative**

---

The Appalachian Regional Commission, or ARC, was established in 1966, with responsibilities that included the expansion of the road, health, and educational systems in what was widely considered one of the nation's poorest regions. Since then, considerable progress has been made in such areas as access to medical care, access to secondary and vocational schools, and elimination of some of the worst effects of previously unregulated strip mining. Despite such improvements, and despite the substantial resources expended by the ARC to improve living and economic conditions in its 13-state area, the economic base of the region retains much of its earlier features. The region's manufacturing base is comprised of firms that are heavily sectorally concentrated in mature industries such as textiles, apparel, food processing, and chemicals. A considerable proportion of employment is in branch plants and non-locally owned manufacturing production facilities. Such ownership patterns may be associated in part with the long-standing policy of many states in the region to attract employment from outside the region to reduce dependence on resource-based industries (such as coal mining).

In the early 1990s, the ARC sponsored a series of sectoral studies of industries in the region. These studies identified serious problems in firm competitiveness. Recommendations included the development of a program to enhance competitiveness by providing technical and strategic

---

information. The Industrial Competitiveness Demonstration Project was launched in response to these recommendations.

The ARC's Industrial Competitiveness Initiative contains two broad project clusters: industry information and worker training. Four projects funded under the Initiative are directed at providing industry information and two projects are directed at worker training. The projects address diverse industries. The industry information projects are directed toward metalworking (including metal forming as well as tooling and machining), plastics (extrusion and injection molding), suppliers to special industry machinery manufacturers, and textile (finishing and cotton) and carpet. The worker training projects are directed toward machine tooling and contractors to chemical manufacturers. The projects differ substantially in terms of their geographic scope, their reliance on cooperating organizations (e.g., on industrial technology agents for identification of clients or assistance in service delivery), and a number of other factors.

The ARC program guidelines require each project to "demonstrate and evaluate the extent to which establishments report improvements in their relative competitive positions (i.e., within their respective industries or markets) as a result of their activities." Additional specific evaluation requirements are specified for the industry information and for the worker training projects.

The ARC realized that six separate project evaluations might not result in either an assessment of the overall worth of the demonstration project, or in learning the lessons available only from cross-site comparisons (e.g., comparative effectiveness of alternative strategies). Moreover, if individual evaluations were conducted with no external impetus toward comparability, it might be extremely difficult to compare across projects after the fact. For these reasons, a cross-site evaluation of the ARC demonstration project was initiated.

---

## Conceptual Framework for Our Evaluation Activities

---

Our conceptualization of possible evaluation activities in the ARC project involves three levels of analysis and four forms of evaluation. Together, these two classifications define the array of potential, alternative evaluation activities.

The ARC evaluation involves three levels of analysis: 1) project level; 2) cluster level; and 3) program level. *Project level analysis*, as the name implies, involves an evaluation focus on the six individual ARC projects. For example, one might ask about the effectiveness of any single project or of each of the six projects. *Cluster level analysis* involves efforts to aggregate findings across discrete substantive groupings of projects. One form of clustering is immediately identifiable in the Initiative announcement: industry information versus worker training. Initial reviews of the six ARC projects suggest the possible emergence of additional clusters (e.g., agent-based versus non-agent-based projects; single state versus multi-state projects). Such clusterings may emerge as having programmatically significant effects in a cross-site evaluation. An example of a cluster-level evaluation question is: What differences exist in the effectiveness/efficiency of interventions across clusters--for example, are agent-based projects more or less effective than non-agent-based projects<sup>1</sup>? *Program level analysis* addresses (if possible) the collective or aggregation of all of the projects in a program, including program level design. An

---

<sup>1</sup>An alternative to focusing on clusters of projects is to focus on project *elements*, that is on specific components of service delivery, such as the type of service delivery provided (e.g., firm assessment vs. technical information). Some of our planned analyses will involve project elements, which can vary within a project, rather than project clusters. Nevertheless, for the sake of simplicity we have not listed both project elements and clusters as distinct levels of analysis.

---

illustrative evaluation question at this level of analysis is: What is the collective impact of the set of projects that constitute the ARC program, on both intended and unanticipated outcomes?

These three levels of analysis arise because the ARC initiative is a multi-site, or more accurately, a multi-project, program. Similar issues about level of analysis can arise in other multi-level programs, such as the Manufacturing Technology Centers (MTC) program.

Whether a program is multi-level or not, multiple forms of evaluation are possible. The domain of evaluation can be categorized in a number of overlapping ways. In the ARC evaluation, the forms of evaluation that we are considering are 1) evaluability assessment; 2) formative evaluation; 3) summative evaluation; and 4) hypothesis testing.

1. *Evaluability assessment* involves determination of the logical connections among a project's definition, set of activities, and presumed outcomes (both short- and long-term) (see, e.g., Rutman, 1980; Wholey, 1987). The traditional purpose of evaluability assessment is to assess whether a project is sufficiently developed to undergo more rigorous evaluation. In the context of the ARC project, evaluability assessment therefore also involves examination of each project's summative evaluation plan, including both specification of appropriate methodologies, measures and management plans (see, e.g., Wholey, 1987). Evaluability assessment can serve a number of functions, including the development of concurrence about outcomes and measures, the development of readiness to utilize evaluation findings, the provision of formative feedback to projects if they lack a coherent or plausible model of the linkages that connect program services to presumed outcomes, and the avoidance of premature summative evaluation.
2. *Formative evaluation* involves monitoring ongoing project activities with a view toward feedback and improvement.

3. *Summative evaluation* involves assessment of overall impact or worth; for example, a project might be judged in terms of whether or not it has achieved its stated objectives or whether it has enhanced competitive position.
4. *Hypothesis testing*<sup>2</sup> involves identification and testing of the hypotheses (explicit and implicit) that arise in the context of summative evaluations of specific projects and across as well as among projects. Both projects and programs are tests of theories. In effect, the six ARC projects represent a variety of more-or-less explicit theories of industrial modernization (e.g., some projects emphasize information delivery systems, and others the validation of management software). The projects also represent implicit tests of alternative delivery strategies for augmenting or replacing the current mainstream technique of reliance on industrial extension agents.

While summative program evaluation seeks primarily to assess the impacts of a given intervention on a target population, hypothesis testing seeks to identify the mediating and moderating relationships that determine the magnitude of an intervention's effects. *Mediating variables* refer to those intermediate linkages through which an intervention has its effects on the outcomes of interest (for example, the effect of Prozac on depression is apparently mediated by changes in certain neurotransmitters). In the ARC project, project effects might be mediated by changes in firm learning. Commonly, mediational models are tested by examining the magnitude and statistical significance of the relationships among the intervention, outcomes, and presumed mediators (e.g., in latent factor structural equation models). More qualitative approaches also exist, such as the

---

<sup>2</sup>In the evaluation literature, this might alternatively be referred to as theory-driven evaluation, testing program theory, or explanation (e.g., Chen and Rossi, 1983; Bickman, 1990; Cronbach, 1982).

---

use of observational and interview data in the context of case studies to probe hypothesized causal links. *Moderator variables* refer to those factors that alter the size of a relationship between two other variables. For example, the impact of providing information may vary depending on whether a firm is large or small, or whether it is a branch plant.

In the case of the ARC projects, the number of potential mediating and moderating variables (e.g., home organization/partnership arrangements; industrial structure), as well as the number of potential confounds, is very large. It is not reasonable to test *all* potential mediators and moderators, for there are many times more potential mediators and moderators than projects. Instead, the focus should be on testing that handful of mediators and moderators that are identified in either of three ways: those that are most prominent in the service-outcome linkages identified in evaluability assessment; those that stand out in site visits and other project review; and those that emerge as important in the continually developing literature on manufacturing modernization. A focus on hypothesis testing should generate useful information for future program revision, and also speak to competing conceptualizations about manufacturing modernization and related issues.

## **Planned Evaluation Activities**

---

At present, we have completed site visits to all of the projects and have reviewed project plans, and have received approval from ARC for a two-stage cross-site evaluation. The first phase of the cross-site evaluation involves activities designed to (a) ensure that the minimum conditions for evaluation, such as the delineation of target populations and project services, are met for all projects, and (b) establish common and/or comparable data elements across projects. These activities are in progress, and they will involve collaboration with project personnel. Consequently, final versions of measures cannot be provided at this time.

The first phase of the evaluation, which may take six months, includes the design and implementation of three data collection instruments, as well as the specification of research designs and identification of control groups. The primary task of the second phase is to carry out a cross-site analysis, including implementation of the design and analysis of quantitative and qualitative methods. In addition, two activities, technical assistance and monitoring, will be carried out through both phases. Some of the activities in the overall evaluation are highlighted below.

### **Project Design Survey**

We have developed a questionnaire, to be completed by each project, that will provide the following baseline information: target population of industries/firms/regions; problem definition; intervention strategy; presumed outcomes; linkages between problem definition, intervention strategy, and presumed outcomes; number of cooperating institutions and form of cooperation. This questionnaire will make explicit information that is important for evaluability assessment, and it will provide a baseline against which subsequent changes can be assessed. Results of this survey, combined with interviews and site visits, will verify whether projects are ready for summative evaluation. Results will also provide project-based hypotheses of the mediating and moderating variables that comprise project theory, and that might be subjected to empirical test.

### **Customer Intake Form**

We are in the process of developing a common intake form with which each project can collect information on the characteristics of firms that use its services. Again, this activity is designed to establish comparability across projects, while also allowing each project to collect information tailored to its particular case. The objective is for each project to have a data base that describes the characteristics of user firms.

The current draft of the survey includes a variety of firm-specific information, specifically name, address, and location of firm, owner/president/CEO education and age, age of firm

---

and of facility, ownership, SIC, major product(s), number of employees, worker skill level, market, distribution channel structure, and how the firm was identified by or made contact with the project. Collaborative planning with projects may result in a revised list of characteristics to be measured.

There are several purposes for collecting this information, including description of the context within which projects are operating, and the assessment of factors that moderate the effectiveness of the services provided (which is critical for theory testing).

### **A Generic “Outcome-oriented” Customer Questionnaire**

Each project's evaluation plan provides for a questionnaire designed to collect information from potential clients. We are in the process of developing a generic questionnaire in order to define a comparable set of outcomes across projects. The questionnaire also will allow each project to insert tailored items.

The survey will be designed primarily to assess the perceived effects, intended and unintended, of project services on firm behavior, operations, and outcome. The work by Oldsman (1994) and the National Institute of Standards and Technology (NIST, 1994) are illustrative of the planned measures of perceived effects. However, we plan to include a greater focus on strategies, policies, and information acquisition. Project services might influence, for example, strategies with respect to marketing, distribution channels, or product lines. These and other changes in strategies, policies and information acquisition may mediate any subsequent effect on sales, jobs, capital expenditures, and savings.

In addition, the current version of our outcome-oriented questionnaire employs a somewhat different format than we have seen elsewhere in the evaluation of manufacturing modernization. Respondents are asked first to indicate the extent to which the firm has experienced change in a particular area, and then to report the extent to which the change is attributable to project services. For the 12-month follow-up survey, for example, one item asks about the level of capital

expenditures in the last 12 months, and the next item asks respondents what percent of the capital expenditures are the result of the project services received.

We believe there are two advantages to such a format. First, we believe it may result in more accurate reports than would arise simply by asking respondents to indicate how much impact they believe has occurred as a result of the services. By thinking first of the total level of change in an outcome (e.g., employment), respondents should make more thoughtful and precise attributions about the impact of project services on that outcome. Second, the items that ask about overall level of change can also be administered to control group respondents, who do not receive project services.

This survey will provide a common metric for cross-site comparisons. It provides key data for summative evaluation and for hypothesis testing.

### **Development and Implementation of a Plan for Cross-site Analyses**

We are developing a plan to summarize the findings from the six projects. One potential product is a comparison across the six projects, and aggregate analyses, of users' perceptions of outcomes. In addition, for those projects for which adequate comparison groups are identified, comparative treatment-control analyses will be attempted to estimate project impacts over and above other sources of change, such as industry trends. Even if treatment-control group contrasts are not available for all projects, those completed will be informative about the level of confidence to be placed in the cross-site comparisons based on user firms' perceptions of effects only. For example, imagine that treatment-control comparisons were available for four of the six projects. Further imagine that in each case, results from the treatment-control comparison suggested a smaller effect than did the treatment group respondents' estimates of effects (that is, when asked to attribute effects to the project services, service recipients estimated larger effects that were observed by comparing treatment and control group firms). Such a pattern would lead one to be cautious about the

---

service recipients' estimates in the two projects without a control group.

The cross-site comparison of user firms' perceptions of outcomes and the treatment-comparison analyses constitute complimentary forms of summative evaluation. In addition, cost-benefit analyses may be conducted, at least for the two training projects.

We also expect to engage in hypothesis testing. The final plan will likely provide for a series of cross-site analyses based on significant policy issues and analytical questions. The policy issues will be identified in consultation with ARC, and will be based on the results of the evaluability assessment. We are currently considering examining the following relationships:

- Processes by which firms incorporate externally provided information or services into routine operations ("firm learning");
- Effectiveness of alternative delivery mechanisms for providing strategic information to firms;
- C Effectiveness of alternative interorganizational relationships; and
- C Characteristics of firms that are associated with greater impact of services.

This list is illustrative and subject to change. Other more important (or more specific) issues may arise.

Various methods may be employed to examine these or other relationships. Often case studies of firms will be employed. Quantitative analyses of possible mediator and moderator variables will likely be conducted.

Previously described activities will also contribute to the methodology of the cross-site analyses. For example, the outcome-oriented customer survey described above can be used to identify "best," "average," and "worst" cases of project impacts, and then to shape subsequent evaluation activities, such as aggregation of firms into "clusters" or the selection of case study sites to facilitate summative evaluation or hypotheses testing. As another example, the customer intake form previously described can provide information on firm characteristics that may moderate the impact of services.

---

## Conclusion

---

A variety of types of evaluation activities can be undertaken for programs such as the ARC Industrial Competitive Initiative. Summative evaluation and hypothesis testing are probably the most important, as they address questions of whether a program or project works (summative), and why or under what conditions (hypothesis testing). However, these evaluation activities may not be warranted without evaluability assessment and, if needed, formative evaluation to ensure that the program/project is "mature" enough for more rigorous evaluation. In addition, evaluability assessment can be extremely valuable in identifying important hypotheses to test, particularly about the linkages that mediate between program services and ultimate outcomes.

The field of manufacturing modernization evaluation should give continued attention to strong research designs for summative evaluation and hypothesis testing. The strongest quasi-experiments and even randomized experiments should be used as possible, though these may be limited to special circumstances. For example, in principle, randomized experiments could be conducted to make the following comparisons: agent-based information transfer versus non-agent-based information transfer; service delivery following an extensive firm assessment versus service delivery without the firm assessment; 60 hours of agent contact available to a firm versus 30 hours or 10 hours of agent contact available. Of course, many other comparisons could be made. The conduct of such experiments would not be without challenges, but the experience of experimental evaluations in many other areas, including medicine, psychotherapy, and criminal justice, indicates it is both possible and worthwhile.

A critical point for future evaluation efforts, not emphasized but implicit in parts of this paper, is that, for good evaluation to occur, program design

---

should facilitate evaluation design. There are a number of alternative program designs that promote evaluation. For example, the planned variation model has been applied in several areas. Alternatively, Campbell (1984) suggests that early on in program development, projects should be allowed to try different approaches and modest standards of evaluation could be applied; then, apparently successful approaches would be disseminated elsewhere and stringent evaluation applied. Other models also exist. What is important is that future developments in evaluating manufacturing modernization may require that modernization programs be designed in such a way as to allow rigorous evaluation. In addition, future work should be fostered by efforts, as embodied in this conference, to create what Campbell (1984) called a "contentious community of scholars," that is, a group of researchers with common interests and interrelated activities whose work and whose reviews stimulate the ongoing development of knowledge.

## Acknowledgements

---

This chapter draws in part on two reports to ARC, "Preliminary Assessment of the Evaluation Strategies of Appalachian Regional Commission Industrial Competitiveness Demonstration Projects" and "A Cross-Site Evaluation of the Appalachian Regional Commission's Industrial Competitiveness Demonstration Projects," (1994), by I. Feller, A. Glasmeier, and M. Mark.

## References

---

- Bickman, L. (Ed.) (1990). *Advances in Program Theory*. San Francisco: Jossey Bass.
- Campbell, D. T. (1984). "Can we be scientific in applied social science?" In R. Connor and others (Eds.), *Evaluation Studies Review Annual, Volume 9*. Beverly Hills, CA: Sage.
- Chen, H. T. and P. H. Rossi (1983). "Evaluation with sense: The theory-driven approach." *Evaluation Review*.
- Cronbach, L. J. (1982). *Designing Evaluations of Educational and Social Programs*. San Francisco: Jossey Bass.
- Feller, I., Glasmeier A., and Mark M. (1994a) "Preliminary Assessment of the Evaluation Strategies of Appalachian Regional Commission Industrial Competitiveness Demonstration Projects" University Park, PA: Pennsylvania State University.
- Feller, I., Glasmeier A., and Mark M. (1994b) "A Cross-Site Evaluation of the Appalachian Regional Commission's Industrial Competitiveness Demonstration Projects," University Park, PA: Pennsylvania State University.
- National Institute of Standards and Technology. (1994). *The MEP Quarterly Report User Guide*. NIST: Gaithersburg, MD.
- Nexus Association (1994). *Evaluation Manufacturing Extension Programs: A Tool for Effective Management*. Report prepared for the Modernization Forum, Cambridge, MA.
- Oldsman, E. S. (1994). "Delivering the Goods: An Evaluation of Manufacturing Extension Programs." Dissertation. Cambridge, MA: Harvard University.
- Rutman, L. (1980). *Planning Useful Evaluations: Evaluability Assessment*. Beverly Hills, CA: Sage.
- Wholey, J. (1987). "Evaluability assessment: Developing program theory." In L. Bickman (Ed.), *Using Program Theory in Evaluation*. San Francisco: Jossey Bass.

---

## Evaluating Technology Institutions and Group Services: Discussion

---

**Jones.** One question is how you ensure appropriate outreach. Should our customers be those most in need or the ones that will have the greatest impact on the economy? In Oregon, we've identified our priority customers, ones that are most important to our economy. These customers tell us that there are too many complex programs. At a time that government is strapped, we're hearing about one-on-one interactions. It would be helpful to learn the best way to deliver services. In Oregon, we feel that helping firms one at a time will not have a significant impact. Networks in Oregon are more focused on significantly affecting the economy. They don't get money for just meeting, they get money for having broader impacts.

**Gillman.** What did you mean when you said that firms in your state thought there were too many assistance programs?

**Jones.** Small companies know what their problems are, and they know that programs to help them solve their problems exist, but it takes too much time to figure out which programs would be appropriate. For example, there are 600 federal programs to assist the forest products industry. In response to companies' complaints about the difficulty of identifying assistance programs, we have reorganized our government departments in response to what companies said they needed. The result is flat departments, organized around customer needs, and headed by team leaders.

**Oldsman.** We found the same thing in New York. We found there were over 435 organizations to which the New York Program had referred its customers. It would be a daunting task for any one firm to keep up with what these organizations do. That should be an extension service role.

**Kaufman.** Pennsylvania is fortunate to have the Ben Franklin Program and the IRCs. The state has made many changes to the Ben Franklin program. The centers were originally tied to the university. Presently, each center is a 501(c)3 non-profit corporation. The state originally established a funding formula based on measures that didn't

provide incentives for the centers to serve the segment of firms in the 25-250 employee range, so a few years ago, we eliminated the funding formula. We allocate equal amounts to each center and award additional funds for innovative activities. The IRCs are being similarly evaluated and changed. We are evaluating each IRC differently because their regional economies are different. We're getting a handle on how to market the benefits of the IRC program to businesses and the legislature. It is misleading to market the IRC program as a job creation program. We've been working on other measures--value-added impact, market penetration, effect of funding level. We believe NIST's and the state's reporting requirements to be burdensome. Why can't they be combined?

**Shapira.** How much should financial measures such as return on program investment (ROI) be emphasized? And, do these measures mean much, since almost every time someone does a calculation, the results are positive. The Japanese don't worry about return on investment. They see their extension activities as a cost of an industrial policy to strengthen manufacturing. We may need to emphasize other measures, for example, improvement in manufacturing effectiveness compared with effort expended.

**Luria.** I am distrustful of trying to do ROI analysis because it encourages comparisons with other programs, for example funding for homeless. It is difficult to do this at the national level, because it creates a zero sum--high returns in some states occur at the expense of the others. The true public role should be seamless. What you really want is an account executive for all manufacturers with 20-150 employees. It sounds utopian, but there are only 70,000 firms that fall into this category. The account executive could connect these manufacturers to good people who would work with them for two days and periodically keep track of these firms' progress. It would take only 1000-1200 field agents to spread that kind of account responsibility throughout the country. Why are we so far from full penetration?

**Russell.** There's a huge price tag to build this kind of robust infrastructure.

---

**Sabel.** I agree with Dan Luria--it's not the resources that are the issue, but how they can be applied effectively. So we first need to figure out what is working and what is not. The Pennsylvania system had to find some way to account for effectiveness. Their situation is similar to that of NIST, so we should pay attention to what they are doing. They've found that the only way to get useful results is to ask questions that the agents feel are useful, that they have defined themselves.

**Kaufman.** Our field agents defined the measure of effectiveness themselves. We call it "value-added" impact. They would like to be able to use it in marketing to a company--for example, if the company makes a certain change, it will have a certain impact. Field agents are hungry for this impact information.

**Oldsman.** Is the role of the field agent to be a catalyst for change or are we saying that many more things have to be done to get a company to change? Oklahoma takes the latter approach. Its staff help companies frame the problems, then manages specialized resources to solve them. Other programs are like the agriculture extension program--agents armed with general information. Which approach works best?

**Mendelowitz.** It is important to look at why firms change. A small number of firms transform themselves because of enlightened leaders. In these cases, NIST centers are helping those who already are aware of the need for change. Others deny there is a problem until they run out of cash, and they either make drastic changes or go out of business. In these cases, NIST centers are assisting businesses that are forced to change. However, NIST centers may have their most important role in helping a different type of company: those that are reasonably well off today, but may not be so in the future. These companies may not have the knowledge or the urgency to seek assistance. Therefore, the best way to reach them is through marketing by the centers.

**Jones.** Have you heard Brian Bosworth's noodle theory? He says that working with industry is like a wet noodle--if you push it, it doesn't move, but if you pull it, it does. Industry cannot be made to participate. Participation has to be on industry's time line.

**Mark.** We can't assume that if we put in a few hours, there will be great change.

**Luria.** If you average the amount of time program staff spend with the firm, it is about two days. I believe that 1000 competent field agents nationwide could make a significant difference in firm attitudes.

**Sabel.** You could reach more companies, but the effect of two days of assistance on those companies is inconclusive. Eric's findings say there is a more arduous range of projects. Which projects those are and how to provide those services are difficult questions. You have to figure out how to provide those services efficiently.

**Oldsman.** My finding says these projects require effort--it doesn't say what kind of effort. The role of the agent might be to assess and understand the problem, and then manage the resources to solve the problem rather than solving the problem himself. We're raising issues that should be built into the evaluation. How many plans include evaluating the referrals?

**Shapira.** Thinking about the number of requests and contacts, is more better or worse? From a pure market economy view, less contacts are better. But, from the view of building an intelligent regional community of firms and service providers, more contacts may be better.