
5

THE VALUE OF EVALUATION: USES, STRENGTHS AND WEAKNESSES

Mark Tebbano

Mark Tebbano is the former Director of the NIST New York Manufacturing Extension Partnership Program (MEP) and serves as Manager of Industrial Technology Programs of the New York State Science and Technology Foundation, the New York MEP's host organization, whose mission is to support technology-based research and business development in New York State.

Manufacturing modernization programs are being established in states across the country largely supported by millions of dollars of public funds. As industrial extension programs unfold, there will be an increasing need to honestly and objectively evaluate their impact and usefulness. However, there are inherent challenges in effectively evaluating a publicly subsidized program whose mission is to assist small manufacturing firms in the private sector to become more productive and profitable.

Profitability is a primary metric in evaluating the success of a private sector venture. Some may say it is the only measure. Increasing the "bottom line" is a goal of owners, investors and managers and employees. Government's "bottom line" is somewhat more subjective and more difficult to measure. Fulfilling some "public good" is a goal of most public programs.

Manufacturing extension programs are usually carried out by non-profit corporations with funding by both government and industry. Impact or success is determined by a combination of "bottom line" and "public good". Such program are designed to provide technical consulting services to customer firms to improve competitiveness and positively impact the bottom line. The public good test is presumably met by companies who continue to operate, provide jobs for workers and pay taxes.

The New York State Science and Technology Foundation, a public corporation affiliated with the State of New York, established the Industrial Technology Extension Service (ITES) in 1985 and expanded it statewide in 1990. Enabling legislation requires that the ITES program be evaluated by an independent consultant every two years to determine the program's impact on manufacturing firms. In 1993 the Foundation engaged Nexus Associates to complete the ITES program evaluation. It represented one of the first such evaluations of a manufacturing extension program. (See Chapter 2 of this volume.)

Value of Evaluation in Manufacturing Extension Programs -- Key Questions to Answer

While perhaps self-evident, it is nevertheless useful to state the importance of effective and acceptable evaluation of manufacturing extension programs. Good information on performance is essential to any organization, particularly new ones in a developing field.

Although there may be general agreement about the value of evaluation, the challenge is carry it out in a public-private environment. We can start by seeking the to answer several key questions:

- C Who should be evaluated?
- C What should be evaluated?
- C When should the evaluation occur?
- C Where should it occur?
- C How much evaluation is necessary?

This workshop undoubtedly will make progress in gaining consensus in answering these questions. The following discussion presumes that an evaluation methodology has been arrived at and generally agreed upon by manufacturing extension professionals.

The Uses of Evaluation

The paramount focus and use of evaluation results should be in determining how to best serve the customer -- the smaller manufacturing establishments. This focus can often be lost in the annual effort to obtain continued funding from federal and state budgets. While evaluation results can be useful in addressing public funding concerns, it is good advice to put the customer first.

Well conceived evaluation results can be used in two ways -- for internal management of the organization, and for external means to improve investor confidence.

Internal Uses

A key use of evaluation results is for the internal organization and management of the

organization providing manufacturing services. Four important areas can be described:

- C First, evaluations can provide assessments on customer satisfaction in terms of how the extension services have impacted the client.
- C Second, the results can be used for continuous improvement of services to customers. Services which have highest benefit can be refined and those with little impact can be eliminated.
- C Third, the information can point to weaknesses in overall program management and administration including data collection, reporting and training of personnel.
- C Fourth, information can be used to assess personnel performance in terms of the quantity and quality of interactions with companies.

External Uses

Effective evaluation results can be used to increase what could be termed "investor confidence" in manufacturing extension efforts. Credible evaluation information is important to public funding decision makers including Congress, federal agencies, state legislatures and executive departments. Similarly, private sources of funding such as corporations and foundations require information on program impact and effectiveness.

Strengths of Evaluation

A concerted, well designed and executed evaluation effort can reap continuous benefits for an extension program. The documentation of results, especially in aggregate form, may provide powerful evidence of program impacts. The information can serve to validate or discredit approaches to extension in terms of services offered and impacts realized. Performance assessments can demonstrate the cost/effectiveness of public fund expenditures.

We have found case studies are one of the most effective means of bringing to life the impact manufacturing extension services can have on firms. Well written case examples often illustrate concepts and services which are otherwise difficult to explain.

Weaknesses of Evaluation

Evaluating manufacturing extension programs is inherently difficult. Obtaining consensus on what the bottom line is and then measuring it continues to be a challenge. Is maximizing client profits the goal? Is productivity? Is employment? Or all of the above?

Honest professional evaluation of extension programs is very complex and time consuming. And accurately quantifying and interpreting results is problematic -- a bit like using a snap shot to tell the story of a whole motion picture.

In the public sector there is the tendency to focus on jobs created and saved through manufacturing extension activities. However, increasing the productivity of firms could lead to reduced employment in the short run. Such results should be handled with care and explanation.

Conclusion

Good progress is being made in moving ahead our understanding about what constitutes "best practices" in tracking the results we seek out of our manufacturing modernization efforts. However, manufacturing extension programs straddle two worlds -- the public and the private. Each have their own measures of success and they don't always coincide. This fact alone should indicate the difficulty we face in professionally evaluating such programs.

Learning from Existing Evaluations: Discussion

Shapira. This is a question for Mark Tebbano. How was Eric Oldman's New York evaluation received by the field staff?

Tebbano. They eventually understood that this was an honest effort to help them. We had an internal task force composed of field staff that guided Eric. His report was widely read.

Oldsman. The study helped define what a field agent is. Now there's interest in establishing a protocol for case management where the agent follows a well-defined process to describe the services provided to a firm. This will help deal with field agent management problems, for example, field agents who like doing quality work diagnosing every customer problem as a quality issue.

Tebbano. We talk about field agents being generalists but that's not true. And such problems as misdiagnoses arise if the field agents aren't managed well. They didn't know what our expectations were. Our new case management process will help.

Heller. Eric could you talk about *post hoc* rather than integrated evaluation? What is the overlap? My sense is there is much overlap.

Oldsman. Your questions bring up two different issues. First there's the need to establish an evaluation system early in the program. New organizations can start from a clean slate, and people are taking advantage of that in formulating their evaluation plans. Second I don't see it as formative versus summative. The evaluation system you would set up at the end asks questions that must be answered throughout the program. That is, what the program's impact is should be addressed all along.

Sabel. Eric's result is the best possible proof you could get that trying to do something matters. More difficult and consequential things will only be done with intense effort. Dan's finding is that what the agent does doesn't matter; it is just the assessment tool. Of course these two findings could be reconciled, because problem identification is

important. Both analyses show that there's a lot to be fixed in the program. For example, in New York there can now be a discussion about whether to offer more easily adoptable services or those requiring more intensive effort. We've seen that we can assess the program, learn, and consider changes.

Tebbano. We all sense that the longer a firm interacts with the program, the greater the likelihood that a change will occur and that the companies will perceive that the agent had an effect on the program. But New York has 28,000 firms--how do we serve them all?

Oldsman. If we can find some low-cost way to get firms over the hump, we can use it. But we don't have the management system to know what those low-cost ways are. One method New York uses that may work is a group project to get companies on the path. We don't know if group projects are truly more cost-effective than one-on-one assistance. Referrals could be another low-cost approach, but there is an effectiveness problem with referrals. In the New York study, thirty percent of the referred firms never called the referral. So one-on-one assistance may be better than poorly managed referrals.

Mark. Chuck Sabel's comments focused on translating evaluation design into program changes. But there are potential conflicts when evaluation is asked to do too much, for example, using program evaluation for personnel evaluation. This could hurt the quality of the program evaluation. If individuals find that program evaluation will affect personnel decisions, these individuals may try to undermine program evaluation.

Malecki. What about customer satisfaction? Customer satisfaction should be a goal.

Oldsman. We looked at customer satisfaction and what factors contribute to it. But there wasn't necessarily a correlation between satisfaction and taking action.

Shapira. But isn't there a problem with bias because most of the responding firms are satisfied?

Oldsman. We are getting mostly satisfied firms and some really irritated, but not many that are somewhat dissatisfied.

Heller. In the Massachusetts design we've invested in the success of performance

benchmarking. The degree to which we can do that depends on whether it is received as a useful tool by the firms involved. But there are practical obstacles we are facing from the field staff. Field staff believe that benchmarking means benchmarking against the best. Are firms in your sample the best?

Luria. The representativeness of the data set is one answer. We call up significant numbers of nonrespondents and haven't found biases in size, industry group, or other category. We solicit nonclients from D&B and Harris data, and we ask trade associations if the distributions look good. Generally we are told that perhaps we're underrepresenting the worst firms, but aside from that it looks good. The comparison groups are custom designed. And the results fit with most firms' experience. Eighty percent of firms believe they are in the top quarter, that is, they think they're better than they are. Seeing themselves compared to their benchmark tends to be a powerful motivator to change.

Russell. I'd add that world class in benchmarking is often a conceit. The choice is usually to close the gap between the company and its competitors, not some abstract notion of what is world class.