

# **NIST/MEP'S EVALUATION ACTIVITY: THE PROJECT TEAM COMPONENT**

**John M. Redman**

**John M. Redman is a Senior Economist with the Program Quality Office of the Manufacturing Extension Partnership at the U.S. Department of Commerce, National Institute of Standards and Technology.**

As a core component of NIST's MEP national evaluation activity, we are establishing nine project teams, each of which will address a specific, high priority evaluation issue. The goal of this activity is to (collectively) define an agenda of specific action items for the MEP's long-term national evaluation plan.

A primary author (or team of authors) will lead each team. The author(s) will draft an initial analysis and recommend specific actions MEP should pursue in that particular issue area. Around the author, we will form a team of an additional 5-7 individuals. The teams will consist largely of NIST, MEP Center and State government personnel, but will also include one or two nationally recognized experts in the topic area.

The reason for extensively including MEP Center and State government personnel on the teams is to make the resulting plan a true collective effort of the entire MEP system, rather than one produced by just NIST headquarters staff. It will also permit systematic incorporation of local, "on the ground" information and perspective, frequently a critically important source of insight. The team members would review the primary author's initial draft, and provide feedback to the author. The author would then revise the draft to reflect these reviews. Additional direct consultations among team members and the author, through, for example, telephone conferencing, will supplement the written feedback and ensure that the revised draft both addresses all issues important to individual team members, and surfaces areas of disagreement.

If necessary, we will then convene each project team, probably at NIST headquarters, around this revised draft. Ideally, each team will reach consensus on the analysis, as well as on specific, high priority action items. We will then integrate these recommendations into an internally consistent long-term evaluation plan. NIST national staff and the Evaluation Working Group will take lead responsibility for synthesizing the individual team products into a coherent action agenda.

We are currently forming, or contemplate forming, nine teams. These nine generally fall into two categories: 1) Data/Methodology Issues and 2) Concept Issues.

---

## Data/Methodology Issues

---

**Benchmarking Panel Data.** We are supporting the Industrial Technology Institute, an affiliate of the Michigan MTC, in developing a firm assessment tool, a database and a questionnaire analysis service which will permit field agents around the country to benchmark a client, or potential client, firm against similar firms on a broad range of performance indicators. While this capability will serve both marketing and firm assessment functions, we also wish to explore its utility in comparing the progress of client firms against that of similar non-client firms, and gauging the relative effectiveness of various types of Center services. Dr. Eric Oldsman will be the lead author for this team.

**Longitudinal Research Datafile/Unemployment Insurance (UI) Databases.** We are contracting with the Center for Economic Studies (CES), U.S. Census Bureau, to examine the feasibility of using its Longitudinal Research Datafile (LRD) in conjunction with the UI database to compare the performance of client with non-client firms, and the relative effectiveness of various types of Center services. The LRD is an establishment-level database, and contains longitudinal data from the Census of Manufacturing and Annual Survey of Manufactures for the 1972-92 period. The UI database contains annual information on employment and payroll.

**Quarterly Report/Client Progress data.** We intend to review the Quarterly Report and Client progress reporting systems to assess their strengths and weaknesses as an evaluation tool. NIST national and MEP Center staff will perform the initial analysis.

**Application of Case Study techniques to program evaluation.** We are contracting with Dr. Robert Yin to help us identify areas in which we might best use case study analysis to support our evaluation activity.

**Program Cost/Benefit analysis.** We are contracting with Dr. Irwin Feller of Penn State for a feasibility analysis of applying cost/benefit analysis to the MEP program.

---

## Concept Issues

---

**Defining Competitiveness.** We will form a project team to revisit the issue of how best to translate the concept of competitiveness into a set of performance measures we can use to track changes in the competitiveness of client firms. Dr. Ariel Pakes of Yale University, Dr. Robert McGuckin of CES and John Redman of NIST/MEP will draft an initial analysis.

**Alternative Service Delivery Mechanisms.** We wish to explore what is known about the relative effectiveness of non-MEP approaches to selecting and delivering services to client firms. This analysis will draw on both U.S. and foreign experience. Dr. Gordon Kingsley of Georgia Tech will lead a team of authors.

**Analysis of Manufacturing Sector Dynamics.** We are also contracting with CES/Census for an analysis of what we know empirically about the linkages among technology adoption, productivity growth, job creation, wage levels, exports and other important indicators of firm performance, and to undertake supplemental analysis to "flesh out" these relationships. This work will, for the first time, use the LRD to develop a comprehensive overview of manufacturing sector dynamics over the last 20 years.

**Coordination of service delivery.** We are contracting with Drs. Philip Shapira and Jan Youtie of Georgia Tech to investigate whether the presence of a center makes an important difference in how well local service providers coordinate the delivery of modernization services to business.

We have limited the number of teams to nine due solely to resource constraints. We recognize that a number of other important issue areas exist, but we have selected these ten as currently of highest priority. However, we sincerely welcome comments regarding this priority selection, as well as recommendations of individuals or organizations who might make exceptional team members.

---

## Evolving Strategies for Evaluation: Discussion

---

**Haines.** One goal of MEP is to understand what aspects of different programs around the nation provide the best services. Another goal is to understand small businesses, how they change, what makes them change. We have started several efforts to fulfill these goals:

First, consistent and routine reporting on a quarterly basis. We have established a uniform set of definitions, and we're requiring every new program to have an evaluation plan.

Second, evaluation infrastructure. We'd like to put together a national plan and influence how programs collect data. We talked to the Under Secretary of Commerce who is concerned about the lack of census data and analysis of small manufacturers.

Third, evaluation working group. Evaluation working groups comprised of volunteers from the larger centers are charged with defining a system for reporting information.

At the national level, MEP is viewed as a government program taking evaluation seriously.

**Kaufman.** You talked about bringing in the stakeholder groups, but you didn't talk about bringing in the states. The states provide matching funds, but are not involved in the evaluation.

**Sabel.** NIST has the capacity to bring together people with many different perspectives. It's very important to pay attention to people who have knowledge. If the people like Terri Kaufman, Janet Jones, and Jan Pounds are not included, the process will be weak. If you embrace state program managers and field staff, they will be the best advocates. But if you impose the "best" evaluation system on them, it will not be the best system. Many evaluators get their best ideas from the people in the field on whom the evaluation system will be imposed.

**Shapira.** How do we operationalize bringing these stakeholders together? You can put them on advisory groups, but is this sufficient? Are there any other approaches?

**Haines.** The Modernization Forum provides opportunities for field staff from different programs to come together and talk, learn from each other, and even work together.

**Russell.** We put together a 300-500 person conference and publish current practice reports based on practitioner information. We expect to have an evaluation tract at our annual conference. But the industrial extension practitioner community shows some resistance to evaluation. We could put together a grouping of senior practitioners, culturally respected by their peers, to be an emissary for evaluation--to help the whole system overcome the negative resistance.

**Oldsman.** Each state program is responsible for developing an evaluation plan that is sensitive to the needs of the program's particular stakeholders. At the same time NIST is developing an evaluation plan for the national program. Which questions are most appropriately answered at the state level? Which questions are most appropriately answered by NIST? How will these two types of evaluation plans be integrated? Is the NIST plan the sum or meta-analysis of these state evaluation plans? It will take time to sort out these issues.

**Redman.** There will always be specific local conditions, but some local issues will be common among all programs, therefore relevant at the national level, whereas other local issues will not be relevant at the national level.

**Oldsman.** One example is that if each state program had a common core evaluation system, if each organization was doing its own evaluation in a rigorous manner, NIST's role would be to synthesize these evaluations rather than to collect and analyze quarterly reporting data.

**Shapira.** Each center should do its own evaluation, to learn about its operations. But NIST has its own customer--Congress, which will want to know things that are not a high priority to individual centers. NIST can also add value by doing things which no individual state can do, such as working with the national data sets.

**S. Martin.** What Eric is suggesting sets a priority for the different levels. For NIST, what is the top priority? Is management of the centers the

---

highest priority? If so, we should be using management-related metrics.

**Sabel.** That is exactly what I was saying. Let's say that all centers provide some common intervention. There should be a common protocol for all common intervention activities. The centers should be able to access and analyze their data themselves.

**Russell.** Everyone wants the focus to be at the center level. The reality is that NIST is bending over backwards to accommodate that universe. But the quality and robustness of the evaluation at the local levels is missing.

**Jones.** NIST can bring quality to this process. NIST can look at the broader picture to help each center be the best it can be. The centers may not be able to have this perspective. The commitment to Congress is similar to how a company would treat its investors. A different kind of reporting is needed for investors. Although NIST serves the needs of Congress, ultimately the centers are NIST's customers, so the measures have to be tailored to the needs of these centers.

**Swanson.** You are struggling to define a new type of evaluation--meta-evaluation. It is a unique form of evaluation I have never heard anyone mention before.

**Heller.** It is reasonable to impose some level of centralization without threatening local autonomy, creativity, and independence. It not enough to require an evaluation plan without requiring specific protocols. Evaluation plans are typically knee-jerk responses to reporting requirements. Reporting requirements don't constitute an evaluation, they constitute monitoring. My experience with federally-funded projects is that program evaluation is less successful when local evaluations exist, and then a national evaluation is imposed as a second system. This places a tremendous burden for providing data. I think meta-analysis and core activities are what we want to move toward. One issue would be to define what a local evaluation plan should include.

**Dugger.** One characteristic of meta-analysis is its flexibility. Another is its ability to create allies. How do you create allies? You have to use a multi-pronged approach. First, you have

experienced evaluators. Second, you need a pre-service element in which center staff are exposed to, and learn about, evaluation. Third, you don't form committees--you leave the process open to everyone who is interested.

**Oldsman.** Centers have to do the quarterly reports, but there's no requirement for evaluation. There's a difference between what constitutes an evaluation plan--it should have a clear statement of objectives, research questions, metrics, procedures, data collection approaches, analysis plan, and means for communicating results--and what is actually needed for meta-analysis. All evaluation plans should have common core elements for data collection and reporting which would sustain a meta-analysis. Why would NIST want meta-analysis? Because NIST can analyze the data without having to collect it. Meta-analysis produces large data sets with which we can detect small changes, which is what we'll probably see in some instances.

**Jones.** I would love to see this as a model for high performance. One way to do this would be to bring in certain trade associations like the American Electronics Association which emphasize benchmarking. We should reward these trade associations for taking the lead. This might change the way trade groups behave, which could then change the way businesses operate. Models in Europe and Asia might have lessons for us in terms of how to involve new partners.

**Kaufman.** Trade associations are a missing link in the national program. They are very involved in school-to-work programs. Their awareness of how to bring the workforce up to the level of the Japanese has led them to take a major role in the accreditation process. There's so many of these programs. Can NIST keep track of them?

**Russell.** NIST is not taking a passive role regarding trade associations. The Modernization Forum will be convening a series of stakeholder councils, the first of which focuses on trade associations. On the issue of meta-evaluations, consider the politics of a locally-intensive, nationally coordinated evaluation scheme. If we have local resistance to NIST's reporting measures now, there will be even more problems if the

---

national program says we have a core evaluation system which must be followed.

**Kaufman.** Everybody says we need core measures. We have core measures--the NIST centers are reporting the same things--but we should allow them to be innovative. Pennsylvania's program works because it is decentralized. I deal with 13 centers--they hate the quarterly reporting that the state puts on them and that NIST puts on them. We have to tell them why they are doing it and what we will do with it. We have to streamline it.

**Heller.** You also have to demonstrate value to them.

**Kaufman.** Will NIST use impact data to evaluate centers? During this start up phase, centers shouldn't be afraid to report bad numbers, because they are learning. We shouldn't compare centers against centers. We should encourage cooperation.

**Oldsman.** In Oklahoma, the first question asked was, what will we do if there are disappointing results? I said I hope there are some disappointing results to learn from. The need to have access to disappointing results emphasizes that the management of the evaluation should be at the local level, so the centers can use the information. But many local plans are sloppy. We have to provide the centers with tools, procedures, skills, competencies.

**Shapira.** In a quality program you don't punish people for reporting bad results.

**Oldsman.** What are people afraid of reporting?

**Shapira.** For example, the result that there is a wash on jobs.

**Russell.** Typically during the second and third trimester, problems emerge.

**Oldsman.** Do these centers think that bad results will be hidden? Even with a rudimentary evaluation system, you wouldn't want to wait 3 years to find out that a center has only dealt with 3 companies.

**Shapira.** When you go on a NIST review, which is only two or three days, you cannot analyze in much depth. Usually the centers are good at producing huge amounts of data--that's part of their strategy. They are presenting their best case to you,

but within the first half day you get a sense of what the data means. All the review panels I've been on have found management problems. Also the lack of center comparative evaluations was highlighted in the first panel's report. That deficiency led to NIST's requiring common reporting measures.

**Swanson.** There are analogs with other programs. One example is the vital registration system. There's an agreed on set of core items, along with variation across the states in terms of what is in the vital registration reports. Reports come out with conclusions such as central Cleveland has infant mortality rates similar to a third world country. People recognize that Cleveland's high infant mortality rates are not the fault of the health program manager.

**Redman.** One center may be operating where unemployment rates are twelve percent, another where unemployment rates are three percent. Just looking at the raw numbers won't tell much nationally.

**Oldsman.** What are the alternatives for control groups? ES-202 files at the state level? There is a limited set of questions that can be addressed with ES-202 data. Census Bureau's Longitudinal Research Datafile? The probability of a firm's being in the Census Bureau's annual survey of manufacturers is small. National benchmarking panels? Who in the company participates in the panel and how are these panels are used? We're working to pilot the use of the national benchmarking panel data, and client progress and satisfaction data that centers have already submitted to NIST. We're going out to the centers to see what data they are actually collecting. There's a lot of data dealing with finance and administration. These data are useful in determining, for example, the value of the contract. We'll be looking at the data inside the centers, pooling it, and making comparisons with Dan's benchmarking control group. One problem is we're not sure how each center is collecting data. Is the wording for determining changes in sales the same in each center? If we wanted to know average costs throughout the program for delivering certain services--a core measure--we couldn't get that because each center is keeping track of costs

---

differently. We can't do comparative analysis without consistent measures.

**Russell.** Standardized questions would add significantly to the robustness of the reports. At the front end, we would want to have a strong dialog with center directors to determine which items have value and what degree of intrusion they would support. We told the centers up front to expect changes in the quarterly reporting, and that we would give them a six-month lead time to make these changes.

**Kaufman.** Your quarterly reporting requirement includes most of the important data. The problem is that each one of the parameters has to be defined. Even when you develop core measures and define them, you still will have people interpreting them differently. Some people won't like the definitions. NIST should be the central agency that says, "these are the core measures." In Pennsylvania, we went through three years figuring out what is a project and what different types of projects we deliver. We developed and have used our value added impact measure. This measure makes it easy for us to sell our legislators on our program. However, this measure is not a catch all. It's just one more impact measure.

**Haines.** We are working on standardization. We have established some working groups to address these issues related to evaluation: training and education, measures/quarterly reporting, evaluation of networks and other alternatives to one-on-one projects, an evaluation plan framework, and uses of evaluation results. We will be providing comparative data from all the centers based on the quarterly reports. A new center director might ask, what is a reasonable number of projects a field agent can work on? Answers to those questions will start to fall out of the data.

**Shapira.** Feeding comparative information to each center may not be very useful. The issue is which centers to benchmark against. Benchmarking against the average is not very useful. You have to know the names of the centers to work out your comparison group. Of course center directors don't like to have their information revealed across the country.

**Jones.** This is a tremendous opportunity to deliver something that will be appreciated by the centers. NIST is getting an overdose of advice. But I'm also concerned about the broader issues about the function of centers, and the future needs and capabilities of partners. My point is that tracking and learning are closely related. There needs to be a focus on helping centers learn and adapt.