

Developing Performance Metrics for Science and Technology Programs: The Case of the Manufacturing Extension Partnership Program

Kenneth P. Voytek is director of research & development with Goodwill Industries International, Inc. Before joining Goodwill, he was chief economist with the Manufacturing Extension Partnership of the National Institute of Standards and Technology (NIST). He earned his master's degree from West Virginia University.

Kenneth P. Voytek
Goodwill Industries International, Inc.

Karen L. Lellock
National Institute of Standards and Technology
Program Office

Mark A. Schmit
National Institute of Standards and Technology

Karen L. Lellock is an economics and performance assessment analyst in the NIST director's office. Before taking this position, she was an economist with the Manufacturing Extension Partnership. She holds an MBA from Mount Saint Mary's College.

Mark A. Schmit is the national account manager for 360vu, the national brand of NIST's Manufacturing Extension Partnership. He holds a master's degree in business administration from Johns Hopkins University.

This article examines the development, implications, and limitations of a series of performance measures to gauge the success of individual Manufacturing Extension Partnership (MEP) centers. The measures represent a shift in how MEP judges and evaluates center performance and a challenge in terms of how they are used, how they are interpreted, and how they are limited. The MEP is an important example of innovative public programs and consists of a public-private, performance-based partnership that seeks to improve the productivity, competitiveness, and technological capabilities of America's manufacturers, particularly small firms. The article makes two important contributions: (a) a comprehensive performance-management approach can be developed with a focus on program outcomes that are linked to long-term impacts and are not just stand-alone process measures or stand-alone outcome measures and (b) it is possible to develop valid and reliable measures for technology-focused economic development programs that can be used to report on and manage performance.

Keywords: *performance measurement; evaluation; manufacturing; technology-based economic development*

Increased attention, reflecting the combination of public interest and political pressure, has brought the issues of performance measurement, program improvement, and accountability front and center for public programs (Osborne & Gaebler, 1992). Historically, public programs focused more on monitoring compliance and performance in terms of inputs and process standards, not outcomes. Recent transformations in public management focus more attention on measuring and monitoring program performance (Hatry, 1999; Kettl, 1997). Measuring, monitoring, and assessing the results of public programs require information that examines not only how much a program spends and on what but also what the public gets in return for the use of these funds and how efficiently and effectively these funds are used.

At the same time, the U.S. economy has undergone a fundamental restructuring as a result of the increasing importance of knowledge, innovation, and technology. Porter (1990) emphasized the

ECONOMIC DEVELOPMENT QUARTERLY, Vol. 18 No. 2, May 2004 174-185

DOI: 10.1177/0891242404263631

© 2004 Sage Publications

importance of the critical link between innovation and competitiveness. In response, many programs aimed at improving science, technology, and innovation among firms have been rolled out to spur economic growth and development. Technology-based economic development programs have grown over time and represent an increasing share of the state and local economic development program portfolio (Feller, 1997; Office of Technology Policy, 2000). These programs range from those providing direct support of basic research to more indirect activities aimed at improving the capacity of firms to innovate and use new technologies. Recently, technology-based economic development policies and programs have focused more on the diffusion and adoption of new knowledge, technologies, and organizational change, thereby raising new challenges in terms of measurement and evaluation (Behn, 1994; Georghiou & Roessner, 2000).

Although the experimentation and innovation are encouraging, they also bring uncertainty and risk. Stakeholders and program managers often have little solid information to help them make policy choices about whether to sink scarce public resources into expanding or starting new initiatives, about what programs are likely to be effective, or about how existing programs might be reshaped to improve performance. The reorientation of programs to a more entrepreneurial, market-based approach suggests that in the absence of a bottom line, the only way to keep score is to measure program performance: the progress made in achieving the goals and objectives of a particular program. The adoption of these new technology-based economic development programs, coupled with demands for increased effectiveness, accountability, and efficiency in spending resources, requires that policy makers and program managers pay more attention to systematically collecting information on program objectives and results. In addition, to the extent programs do measure performance, the focus is usually on inputs and outputs, not program outcomes and service quality. The absence of such information is becoming more serious as policy makers face mounting political pressure to revitalize their economies on one hand and face increasingly tight fiscal resources to support public programs on the other hand.

This article outlines and examines the Manufacturing Extension Partnership (MEP) program and the methodology underlying the development of center-specific performance measures that shift from a focus on program inputs and outputs to a focus on measures of program outcomes. It examines the transition to a more comprehensive view of performance, explores the challenges posed, and illustrates that a series of performance measures can be developed that provide managers and policy makers with the information needed to improve the design, management, and execution of these programs, as well as provide information on the efficiency and effectiveness of individual centers. The article not only shows that programs must have the capacity to measure performance but also that program managers must use this information. It also adds to the literature on performance measurement, particularly those programs focusing on technology-based economic development (Hatry, Fall, Singer, & Liner, 1990).

THE MEP PROGRAM

Created in 1988, the MEP program seeks to improve the productivity, economic competitiveness, and technological capabilities of America's manufacturers, particularly small manufacturers. MEP is a results-based partnership consisting of a network of locally operated, staffed, and controlled nonprofit and university-based organizations leveraging federal, state, local, and private resources. The program consists of 60 manufacturing extension centers throughout the United States and Puerto Rico. The size of each center was established largely based on its ability to match federal funding at the time of the initial proposal and does not reflect a distribution based on the concentration of industry in its service region. Centers are funded by a combination of federal, state, and local contributions and client fees generated from the services provided. Twenty-eight centers receive less than \$1 million in federal funding, 17 centers receive between \$1 million and \$1.9 million, and 15 centers receive \$2 million or more in federal funding. The largest MEP center received 45 times as much federal funding as the smallest center in fiscal year 2000.

Each center works directly with local firms to provide expertise and services tailored to their most critical needs, ranging from process improvements and employee training to new business

Recently, technology-based economic development policies and programs have focused more on the diffusion and adoption of new knowledge, technologies, and organizational change, thereby raising new challenges in terms of measurement and evaluation.

practices and the application of information technology in their companies. The program focuses on the diffusion and adoption of new knowledge and technology among small manufacturers and includes working with these firms to develop or adopt new products and processes, incremental improvements to products and processes, and new approaches to marketing and distribution. Services to client firms are delivered through either direct assistance from center staff, assistance from outside partners, or a combination of the two.

MEP'S PERFORMANCE-MEASUREMENT STRATEGY

Performance measurement and evaluation are the cornerstone of the MEP program. Performance-measurement systems and a comprehensive strategy were built into the program as it began. The strategy has evolved over time. The program devotes considerable resources to measuring and monitoring center performance, including:

- Quarterly data reporting. Each center reports data on a variety of areas quarterly. Major elements of quarterly reporting include progress data, progress narrative, and an Activity Data Log describing projects with clients and other information.
- Operating plans. Each center prepares an annual operating plan outlining its proposed activities and results for the coming year. The operating plan is reviewed and approved by MEP and forms the basis for monitoring progress throughout the year.
- Annual reviews. Before being renewed, each center must be reviewed by MEP, and the results of that review must recommend continued funding. The annual review is modeled on the panel review process and uses the *MEP Criteria for Center Performance Excellence* (National Institute of Standards and Technology, Manufacturing Extension Partnership, 1998). A report is prepared after each review outlining a series of recommendations that are part of the center renewal package.
- External panel reviews. The MEP statute (The Omnibus Trade and Competitiveness Act of 1998) requires an external panel review of each center during Years 3 and 6 and every 2 years thereafter. Review panels use the *MEP Criteria for Center Performance Excellence*, based on the Baldrige criteria. A center prepares a Center Progress Report, outlining the key aspects of its operating model, strategy, and results. Center staff and the panelists discuss the report, and the panel develops recommendations.

Early on, performance measurement in MEP concentrated principally on internal organizational processes and administration and program inputs and outputs. As the program and centers started up, the focus on these aspects was warranted, but as the program has matured, some questioned whether the focus should shift. In early 1999, MEP developed a series of indicators that, although an improvement, were not clearly tied to the review process and focused primarily on program inputs and outputs. In addition, although Section 7.1 of the *MEP Criteria for Center Performance Excellence* highlights the importance of reporting results, the data used by centers to measure and report on results were ad hoc and did not allow for consistent comparison across time and across centers. The National Research Council's (1993) report, *Learning to Change*, observed that "coherent measures . . . should be developed for evaluation of federal, regional, and local assistance efforts. . . . A positive evaluation of assistance organizations should reflect . . . success by their manufacturing clients" (pp. 89-90). Moreover, the MEP National Advisory Board's Subcommittee on the Evaluation Process recommended "that MEP rebalance its portfolio of evaluation frameworks . . . shifting increasingly to results-oriented measures" and urged that the program concentrate its future efforts on developing reliable and valid outcome measures (National Institute of Standards and Technology, Manufacturing Extension Partnership, 1999, p. 20). MEP's senior management also recognized that the program needed new ways to measure and judge center performance.

. . . . the MEP National Advisory Board's Subcommittee on the Evaluation Process recommended "that MEP rebalance its portfolio of evaluation frameworks . . . shifting increasingly to results-oriented measures" and urged that the program concentrate its future efforts on developing reliable and valid outcome measures.

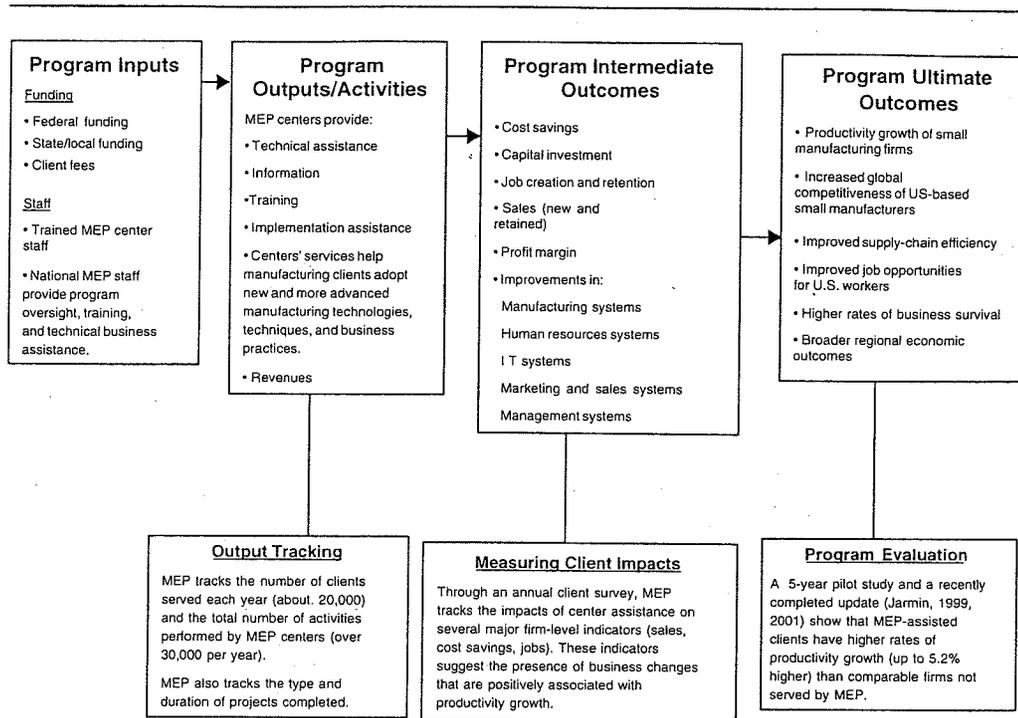


Figure 1: The Manufacturing Extension Partnership (MEP) Program Impact Path

THE PROGRAM MODEL

Few argue against the need for performance measures, but there is considerable debate over what should be measured and how (Feller, 2002). The development of performance metrics for technology-based economic development programs is challenging because the link between program inputs, outputs, and outcomes (the so-called production function) is uncertain and loosely coupled and programs may have multiple goals that may not yield measurable short-term outcomes (Behn, 1994; Feller, 2002). An important challenge in developing a performance-measurement system is to understand the program's production function or logic model (Scheirer, 2000; Wholey, 1977).

The use of logic models provides information on the conceptual connections between program inputs, outputs, and outcomes and what the program measures along this path that reflects the translation of inputs into outputs and finally outcomes. In addition, significant differences across the centers in terms of program inputs and outputs and program models presented important challenges to performance measurement. The model helps provide a basis for understanding how the various measures could be pieced together to explain why performance varies across the centers. Such a model also helps show how two important purposes of performance measurement—program improvement and public accountability—can be linked theoretically and in practice (Scheirer, 2000).

Figure 1 highlights, in a stylized fashion, how the MEP program operates in terms of translating program inputs into outputs and ultimately into intermediate and long-term (or ultimate) program outcomes in the upper band. The lower band provides information on how MEP measures and monitors performance at various points along the way. The focus of this article is on the metrics that measure intermediate program outcomes associated, both conceptually and practically, with productivity improvements, which is the key long-term program objective (Voytek, 2001). These intermediate measures are strongly correlated with the program's long-term goals, and the weights

are derived to ensure that no one particular measure predominates, which could result in perverse incentives or unintended consequences (Heinrich, 1999).

MEASURING PROGRAM IMPACT: CONCEPTUAL CONSIDERATIONS, THE DATA, AND THE METRICS

.... the measures were created to inform both policy and program-management decisions and were designed to be generated regularly and publicly with an emphasis on clarity, transparency, and timeliness. Significant time was devoted in helping the centers understand how the measures were calculated, what they were, and how they would be used.

With a well-developed system in place to collect information on the program's major components, the next step was to develop a new generation of performance measures. The goal in developing these performance measures was to provide regular and valid information on program outcomes. Five measures were developed in collaboration with MEP senior management and program managers, the MEP National Advisory Board, members of the Center Review Panels, center staff, and other experts.

Several important considerations guided development of these measures. First, the focus was on client outcomes, with a few key indicators measuring client impacts. Second, a consistent set of measures over time and across centers was generated. Third, contextual information was used to help understand variations in performance over time and across centers to facilitate continuous improvement and program accountability (Newcomer, 1997). Fourth, the measures were created to inform both policy and program-management decisions and were designed to be generated regularly and publicly with an emphasis on clarity, transparency, and timeliness. Significant time was devoted in helping the centers understand how the measures were calculated, what they were, and how they would be used. The development of the measures was intended to be as transparent as possible. Center performance data, including the raw data, were provided to the centers, and they were soon to have access to information on the performance of all centers. Fifth, the unit of analysis was the 60 centers composing the MEP network.

Since 1996, MEP has conducted a national survey to gather information on the client-level effects of services provided by local manufacturing extension centers. The survey is conducted four times each year and consists of 20 questions focusing on the following:

- bottom-line client outcomes, such as sales, technology, capital investment, cost savings, and productivity improvements;
- intermediate process improvements in areas such as manufacturing operations, sales and marketing, human resources and training, and information and management systems; and
- client satisfaction with the services provided.

Although MEP had measured and reported on the program's aggregate impact, it had not systematically used the data to judge and assess center-specific performance. The aggregate performance measures are useful for overall program management, but they were of little use in assessing variations in performance across centers or understanding how these results were achieved (Affholter, 1994). The aggregate program performance data masked substantial variation among the individual centers in terms of their impact.

At the end of each survey, the client-level data are aggregated for each center, and the measures are calculated for the most recent four quarters. A rolling four-quarter aggregate is used to examine center performance. A four-quarter period was selected to ensure that the measures are not

- influenced by a single survey period (substantial variation across centers reflects differences in their respective fiscal years, which influence when projects are closed. Thus, there are variations in the number of surveys done quarter to quarter);
- dependent on a single quarter's results because a center's performance varies from quarter to quarter and the intent is to focus on long- rather than short-term performance; or
- skewed by double counting client impacts because clients are only surveyed once a year.

The survey is client-based rather than project-based. Clients completing multiple projects with a center are surveyed only once per year. Clients with multiple completed projects are asked to consider the entire set of projects or services provided by the center and report how their company's performance and processes have been affected in the past 12 months. The survey is also retrospective rather than prospective. Clients are asked to report on the realized, rather than the expected, benefits of the work completed.¹ This survey approach minimizes client burden and provides an unduplicated count of clients in any four-quarter period. Clients may complete the survey using one of three modes: telephone interview (a vast majority of the surveys are completed this way), interactive voice response system, or Web-based survey. In fiscal year 2001, 6,113 clients were selected for survey, and 4,804 clients completed the survey, resulting in an overall response rate of 78.6%.²

The measures were developed from the perspective of MEP as an investor in a center. They consider client-level impacts in terms of the gross return on the federal investment in a center. The variation in funding levels across the centers noted earlier suggests that an assessment of performance should reflect the relative level of resources available to each center. Using federal investment in the calculations normalizes the data, allowing for comparisons across centers and over time. Additional performance categories were developed to discriminate more discretely among centers. Centers do not simply pass or fail. Centers earn a score that varies along a continuum. Finally, consequences are attached to performance. Centers failing to achieve the minimum expected performance levels face increased scrutiny if performance does not improve.

These measures were also developed to ensure that they did not dictate or mandate a particular mix of services or a service-delivery approach. A center's service mix and approach should be determined based on its clients' needs and should be supported by internal staff, augmented as needed with the expertise of its partners and third-party providers.³

The five performance measures⁴ are as follow:

- Cost per impacted client. The federal investment in a center divided by the total number of clients responding to the survey who reported some business impact, such as an improvement in their profit margin, sales impact, cost savings, productivity gains, jobs impact, or new investment. This measure provides an indicator of a center's ability to efficiently generate client impacts.
- Bottom-line client-impact ratio. The sum of cost savings reported by clients plus 15% of total sales impact (new sales plus retained sales) divided by the federal investment in a center. This measure was designed as a proxy for the bottom-line impact of services delivered by a center. Only 15% of the sales impacts are used to reflect an estimate of the gross margins for all manufacturers based on an analysis of data from the *Annual Survey of Manufacturers* published by the U.S. Census Bureau (2000). This measure provides an indication of a center's ability to help its client firms develop and retain resources and revenues.
- Investment leverage ratio. The sum of new investment reported by clients (defined as new investment in plant and equipment, information systems and software, training, and other areas) divided by the federal investment in a center. Investing in human and physical capital is an important ingredient in improving the productivity and competitiveness of small manufacturers. Investment leverage ratios are a commonly used performance metric in many economic development programs.
- Customer satisfaction score. The average customer satisfaction score reported by clients responding to the survey. Clients are asked to rate their satisfaction with services provided on a scale of 1 (*very dissatisfied*) to 5 (*very satisfied*). Responses are summed across all clients and divided by the number of clients responding. This measure has the advantage of providing a simple numerical summary but does not indicate how the responses were distributed. The rank order correlation among the centers in terms of customer satisfaction score and the tails of the distribution (i.e., the percentage of clients reporting that they were very or somewhat satisfied and very or somewhat dissatisfied) are positive and statistically significant.

These measures were also developed to ensure that they did not dictate or mandate a particular mix of services or a service-delivery approach. A center's service mix and approach should be determined based on its clients' needs and should be supported by internal staff.

TABLE 1
Initial and Current Weights of
Manufacturing Extension Partnership Performance Measures

	\$ Cost Per Impacted Client	Weight in Points	Bottom- Line Client- Impact Ratio	Weight in Points	Investment Leverage Ratio	Weight in Points	Customer Satisfaction	Score Weight in Points	% Survey Response Rate	Weight in Points
Old threshold	43,500	10	2.00	30	2.30	30	4.1	15	60	15
New threshold	31,000	30	3.00	30	3.00	20	4.2	10	70	10

- Survey response rate.⁵ The number of clients completing the survey divided by the number of clients selected. The survey response rate, in conjunction with the customer satisfaction score, provides a good overall measure of customer satisfaction. The survey response rate also provides a measure of the quality of the survey responses in terms of nonresponse bias.

The metrics are now widely accepted by the centers and others as a way to level the field, put performance assessment on an objective basis, and communicate to centers what is important to one of their primary partners—their federal investor. Nearly 3 years of performance data are now available, allowing an analysis of performance trends over time.

A center is scored on each measure, and individual scores are summed across these measures to determine an overall score for a center. At the end of each survey period, centers receive a report detailing their performance on each measure for the most recent four-quarter period, their overall aggregate score, and information on their relative performance on each measure and overall in terms of the median and various percentile rankings (i.e., 10th, 25th, 75th, and 90th). Centers are often interested in knowing how they compare to their peers and have begun to benchmark those that consistently achieve high rankings.

KEEPING SCORE: WHO'S ON FIRST?

A center's score can range from 0 (below the minimally acceptable level of performance on all five measures) to 100 (performance at or above the minimum level on all five measures). An aggregate score of 70 or more is needed to be considered in good standing with MEP. A center receives points for a particular measure if it is at or above the minimum performance level established; points are summed across the five measures to get a center's final score.

Initially, more weight was placed on the bottom-line client impact and investment leverage ratio than on the other measures. Table 1 details the initial performance targets and weights for each measure and those currently in place. The weight on the cost per impacted client was increased because it provides a measure of market penetration superior to simple counts of clients served. In addition, the idea of weighting two measures—cost per impacted client and bottom-line client impact ratio—at 30 points each was to highlight the trade-off and strategic choices that centers face.⁶

Initial performance targets were determined by examining the data for all centers over a four-quarter period. The initial minimally acceptable performance levels were set at about the 25th percentile of all the centers for each measure. The new weights and thresholds reflected a desire to continue to ratchet up performance targets and reflect the importance of both impact and market penetration as program goals.⁷

Table 2 provides data on each of the measures at two different points in time: the first four quarters in which the data were collected and used (Q1–Q4 2000) and the most recent four-quarter period covering surveys done between third quarter 2001 and second quarter 2002. Data on the distribution of the metrics across the centers, including the 25th, the median, and 75th percentiles, are

TABLE 2
Initial and Current Performance Levels
Across the Manufacturing Extension Partnership Centers

<i>Survey Period</i>	<i>\$ Cost Per Impacted Client</i>	<i>Bottom-Line Client-Impact Ratio</i>	<i>Investment Leverage Ratio</i>	<i>Customer Satisfaction Score</i>	<i>% Survey Response Rate</i>
Quarter 1 2000 to Quarter 4 2000					
25th percentile	43,747	1.97	2.32	4.11	58.9
Median	23,914	5.83	5.17	4.28	64.9
75th percentile	15,964	13.96	12.48	4.39	72.1
Quarter 3 2001 to Quarter 2 2002					
25th percentile	33,334	4.83	3.10	4.34	73.9
Median	20,270	8.11	6.06	4.45	80.0
75th percentile	14,133	15.13	10.99	4.52	85.7

presented to show how performance has changed over time. Center-level performance has improved, and the gap between high performers and low performers has narrowed but remains substantial. For example, the bottom-line client impact ratio was more than 7 times higher than the 25th percentile in 2000 but only 3 times higher in the most recent four-quarter period. Further analysis at various break points suggests that the low-performing centers need to look beyond increasing capacity utilization and service-delivery efficiency (certainly in terms of managing the response rate) because the performance gaps along these dimensions are much narrower and do not reflect the observed differences in inputs and outputs. Centers must look more closely at the strategic choices they make, including the types of clients, the types of services delivered, and the way the work is done.

MOVING FROM DESCRIPTION TO DIAGNOSIS

To focus on the strategic choices centers face in terms of working with lots of clients versus creating large impacts, the program developed a Metrics Map.⁸ This map plots the performance of centers on the two most important dimensions of center performance—cost per impacted client and bottom-line client-impact ratio. All centers can be arrayed in terms of their success along these dimensions at a particular point in time. Figure 2 shows relative performance across the centers. The dots represent centers, and there are two dots for the entire program (labeled MEP and Ntnl). The lines (labeled 3M, 5M, etc.) on the chart are intended to represent crude returns on investment and represent 3-, 5-, 7-, 10- and 20-to-1 rates of return. Centers can be compared to others, and their performance can be examined over time. Improvement represents movement in a northeasterly direction from the origin. The x-axis transforms a center's cost per impacted client metric and has the advantage in that moving out from the origin and along the x-axis shows improvement. It also normalizes the data by assuming that each center receives \$1 million in federal funding. To derive the x-axis, divide \$1 million by a center's cost per impacted client. For example, say that Center A's cost per impacted client was \$20,000. The coordinate on the x-axis for Center A would be 50 clients. The y-axis measures impact intensity. Moving up the axis represents an improvement. Multiplying the cost per impacted client by the bottom-line client impact ratio derives a position for a center on the y-axis. For example, say that Center A's bottom-line client-impact ratio was 10. The coordinate for the y-axis would then be \$200,000 ($10 \times \$20,000$). Thus, Center A's dot would be placed at the intersection of these two points.

The intent of this chart is to move centers and the review panels from a focus on operational issues to a focus on strategy and choice. This chart moves the discussion to this level by displaying the trade-offs that need deeper understanding and wider discussion. The chart helps centers think

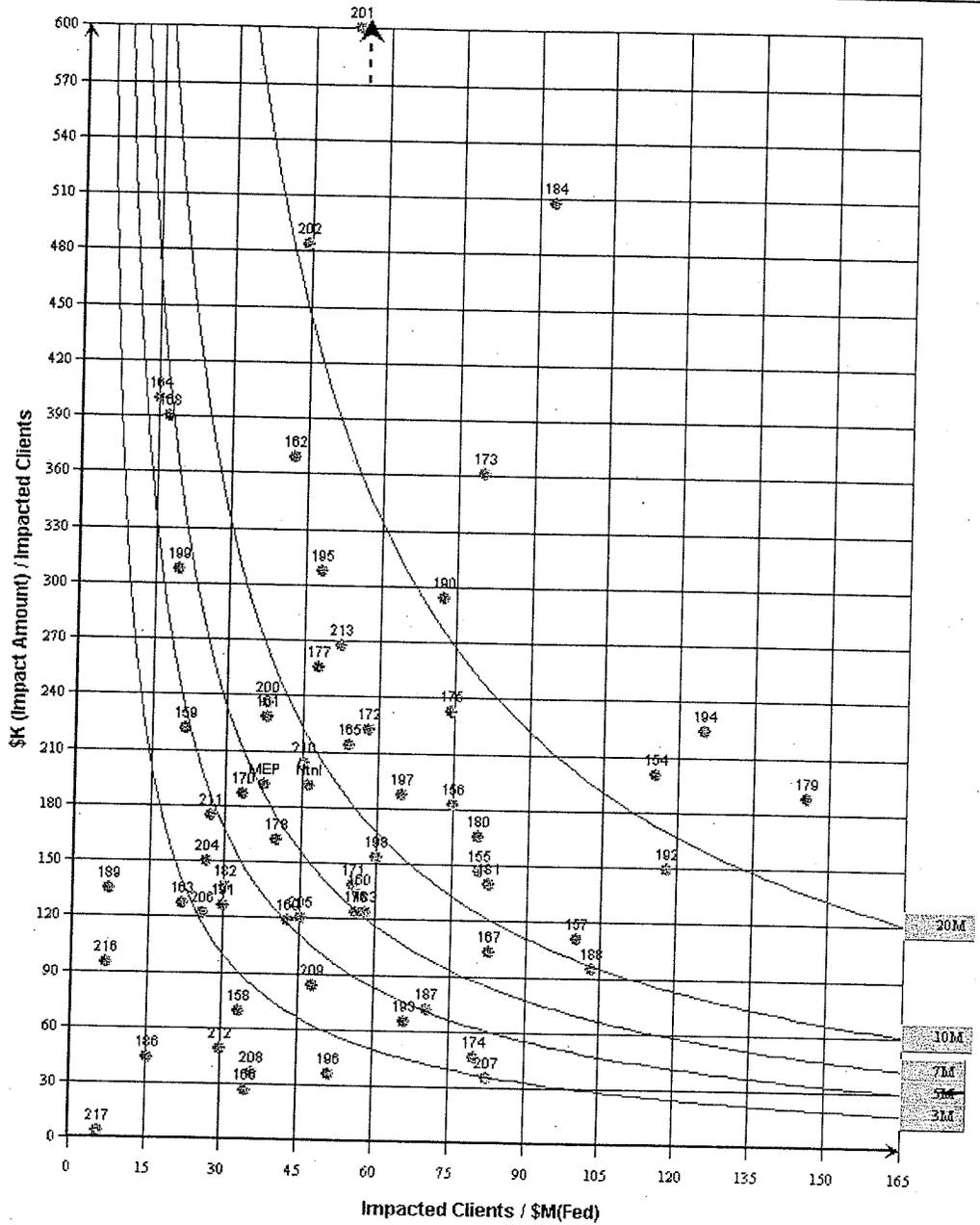


Figure 2: The Manufacturing Extension Partnership (MEP) Metrics Map, Period Ending Quarter 1 2002

NOTE: \$M = dollars in millions. Fed = federal funding. \$K = dollars in thousands. Each dot with a number represents a specific center. The lines labeled 3M, 5M, 7M, 10M, and 20M represent return on investment ratios. The dot labeled MEP represents the return on MEP program with both the federal investment in centers and headquarters functions included, and the dot labeled Ntnl represents the return on investment if only the portion of federal funding going to centers is included.

The map also allows centers to compare and benchmark their performance on these two dimensions more readily. For example, Center 207 could benchmark and investigate what other centers such as Centers 167, 161, 155, or 180—with a similar number of impacted clients but with higher average impacts—are doing. In addition, given that these measures are available for several time points, a center's performance or movement over time can be analyzed.

about what projects they deliver, how they deliver, and whom they work with. The x-axis relates to penetration intensity and largely reflects differences in the service delivery approach that centers

adopt, as well as project size. Centers can deliver services through third parties and serve as project managers, or they can deliver services using their own staff. They can also work with clients on shorter duration projects. Centers closer to the origin on the x-axis are more likely to use their own staff to deliver work. Centers farther out on the x-axis are more likely to use a broker model that involves the use of third-party organizations to deliver services. The y-axis reflects impact intensity and differences in the types of projects delivered, project length, and the types of firms worked with. The delivery mode should not, in theory, be strongly linked to impact intensity unless the choice of direct versus broker models results in very different impacts for the same services and firms.

The performance gaps are, based on an analysis of efficiency and capacity utilization rates, much too large to be closed by incremental and marginal operational improvements. The dramatic differences in performance do not reflect differences in efficiency or capacity utilization because they are narrower than the performance gaps.

Displaying outcomes this way also represents a new approach to the discussion of center performance. It presents important challenges to better understanding centers and to assembling data and information for making better diagnoses and recommendations. In fact, it is this type of analysis that helps focus the program and centers on the delicate balance that performance measures are used for—ensuring program accountability but also to spur continuous improvement. It also moves the discussion of continuous improvement to a higher level in that it is more likely to lead to important and nonincremental performance improvements.

LIMITATIONS AND NEXT STEPS

The development, use, and deployment of these measures present new challenges. First, these measures are not intended to be the end of a discussion. Measuring performance is not the same as managing performance. The measures are intended to begin a discussion with a center in terms of how it has performed, how it has changed over time, and what it intends to do to improve its performance. This requires a higher level of sophistication in how the measures are used by program managers, review panels, and center staff. More training on how they can be used and interpreted is needed.

Second, the metrics do not, by themselves, provide information on how the outcomes were achieved or why a center performed at a particular level. Some internal and unpublished work we did attempted to explain the variations in performance, and the results were suggestive. The analysis found that variations in service-delivery approaches (centers employing a broker model did "better"), variation in state funding and particularly the "quality" of that funding (the portion of state funding that was cash rather than in-kind was associated with improved performance), and variation in the types of projects delivered by a center (measured in terms of the split of a center's work that were technical assistance and training projects) were all strongly and positively related to center performance. Several other factors such as center age, client density, federal funding levels, and revenue generation, did not appear to be related to variations in center performance. More work needs to be done in terms of understanding what factors account for variations in center performance both to assist in diagnosis and to inform center decision making. Early work using Data Envelopment Analysis was undertaken by MEP, but this analysis was done before the introduction of these performance measures and needs to be extended (Chapman, 1998).

Third, these performance measures are about the past. Clearly, the data collected are backward-looking, and improvements in center performance may take time to emerge. However, given that the measures are based on clients rather than particular projects, the outcomes reported may measure some of these changes in center operations. It may, however, be important to develop some early warning indicators that provide some clues as to the possible future direction, but they cannot be used to definitively estimate or predict outcomes in the future.

Fourth, the metrics do a good job in identifying the top and bottom performers, but they provide little information to help discriminate or deal with the so-called muddle in the middle. The performance measures are a blunt tool and cannot discriminate among centers in precise ways. In addi-

The measures are intended to begin a discussion with a center in terms of how it has performed, how it has changed over time, and what it intends to do to improve its performance. This requires a higher level of sophistication in how the measures are used by program managers, review panels, and center staff.

tion, the measures are subject to some error and can be influenced by a large impact. Validating large impacts is done systematically, but large impacts do influence a center's performance. Some have suggested that the data should be trimmed, and how to do this, beyond simple statistical techniques, is unclear. The use of a rolling four-quarter time frame smoothes out some of the swings, but even this measure is subject to the influence of a single large impact.

Fifth, the metrics do not reward good performers or performance improvements. Currently, the primary use of the metrics is to improve center accountability in terms of focusing on minimum levels of performance, but the measures do not provide incentives or rewards for centers to improve their performance. These measures, by themselves, may not be appropriate for making resource allocation decisions alone. In addition, the measures focus only on outcomes. Work is under way to develop other measures beyond the outcome measures examined here.

CONCLUSION

Innovation and knowledge are important not only to economic performance but to program management and policy decisions. An important step in furthering the practice of economic development is learning about and sharing best practices and innovative approaches. More attention to devising performance measures is an important step in improving the management and execution of economic development programs and policies. But they also present management challenges in how they can be used to spur continuous improvement and provide information that is actionable on the part of program managers or center staff.

This article provides an example of how performance measures can be developed and applied to achieve the following two goals: increased program accountability and continuous improvement. The article explored how a family of performance measures can be developed that can inform both management and policy decisions to improve the design, management, and execution of a program. The methodology and theoretical considerations that guided the development of center-specific performance metrics were examined. This article also examined the key issues revolving around the development, deployment, implications, and shortcomings of the metrics to gauge the success of individual MEP centers. This article extends the program monitoring, performance measurement, and evaluation capacities of program managers, particularly those responsible for technology-based economic development. It also shows that these tools could be applied to other economic development programs and even other public programs.

NOTES

1. Neither approach is without its drawbacks. Both approaches introduce bias into the estimates of the effects of projects completed. A limited analysis of data from centers that do project close-out surveys suggests that the aggregate impacts do not differ markedly but the individual impacts do. In addition, the retrospective approach does present problems in terms of recall on the part of clients and backward-looking measures.

2. Additional information on the survey methods, the instrument, and the steps taken to minimize both response and nonresponse bias to the survey, as well as an examination of the characteristics of the clients who responded to the survey and those who did not to explore whether there were indications of potential response bias, is available in National Institute of Standards and Technology, Manufacturing Extension Partnership (2002).

3. Other measures focusing on productivity, efficiency, and capacity utilization have been developed but are not the focus of this article.

4. Noticeably absent from this list is a jobs measure. The survey collects data on jobs retained and created by clients, but this information is not used because the program has placed less emphasis on this particular dimension. Data are reported to the centers on the cost per job created and retained, but they are not part of the family of impact measures currently used by the Manufacturing Extension Partnership (MEP).

5. The survey response rate, although not a direct measure of impact, was included to help centers understand that their initial emphasis should be on getting clients to respond to the survey.

6. The trade-off is not as linear as implied, nor does working with clients on longer projects necessarily lead to larger impacts or vice versa. However, it is important for program managers and center staff to recognize this choice.

7. Every March, the program staff reviews the measures and the weights and adjusts them as needed.

8. Dr. Curt Reiman, a consultant to the MEP, developed the initial map, and we would like to acknowledge his important contributions.

REFERENCES

- Affholter, D. (1994). Outcome monitoring. In J. S. Wholey, H. P. Hatry, & K. Newcomer (Eds.), *Handbook of practical program evaluation* (pp. 96-118). San Francisco: Jossey-Bass.
- Behn, R. D. (1994). Here comes performance assessment—and it might even be good for you. In A. Teich, S. Nelson, & C. McEnaney (Eds.), *Science and technology policy yearbook—1994* (pp. 257-264). Washington, DC: American Association for the Advancement of Science.
- Chapman, R. E. (1998). *Using Data Envelopment Analysis to assess performance of Manufacturing Extension Centers*. Gaithersburg, MD: National Institute of Standards and Technology, Office of Applied Economics.
- Feller, I. (1997). Federal and state government roles in science and technology. *Economic Development Quarterly*, 11(4), 283-295.
- Feller, I. (2002). Performance measurement redux. *American Journal of Evaluation*, 23(4), 435-452.
- Georgioui, L., & Roessner, D. (2000). Evaluating technology programs: Tools and methods. *Research Policy*, 29(4-5), 657-678.
- Hatry, H. P. (1999). *Performance measurement: Getting results*. Washington, DC: Urban Institute Press.
- Hatry, H. P., Fall, M., Singer, T. O., & Liner, E. B. (1990). *Monitoring the outcomes of economic development programs*. Washington, DC: Urban Institute Press.
- Heinrich, C. J. (1999). Do government bureaucrats make effective use of performance management information? *Journal of Public Administration and Theory*, 9(3), 363-393.
- Jarmin, R. S. (1999). Evaluating the impact of manufacturing extension on productivity growth. *Journal of Policy Analysis and Management*, 18, 99-119.
- Jarmin, R. S. (2001). *The impact of the MEP on plant productivity* (Draft paper). Washington, DC: U.S. Census Bureau, Center for Economic Studies.
- Kettl, D. F. (1997). The global revolution in public management: Driving themes, missing links. *Journal of Policy Analysis and Management*, 16(3), 446-462.
- National Institute of Standards and Technology, Manufacturing Extension Partnership. (1998). *MEP criteria for performance excellence*. Gaithersburg, MD: Author.
- National Institute of Standards and Technology, Manufacturing Extension Partnership. (1999). *Report to Congress—The NIST Manufacturing Extension Partnership: A network for success and a review of the results and evaluation process*. Gaithersburg, MD: Author.
- National Institute of Standards and Technology, Manufacturing Extension Partnership. (2002). *The Manufacturing Extension Partnership: Delivering measurable returns to its clients*. Gaithersburg, MD: Author.
- National Research Council. (1993). *Learning to change: Opportunities to improve the performance of smaller manufacturers*. Washington, DC: National Academy Press.
- Newcomer, K. (1997). Using performance measurement to improve programs. In K. Newcomer (Ed.), *Using performance measurement to improve public and nonprofit programs* (pp. 5-14). San Francisco: Jossey-Bass.
- Office of Technology Policy, U.S. Department of Commerce. (2000). *Federal role in support of technology-based economic development: Report of a roundtable*. Washington, DC: Author.
- The Omnibus Trade and Competitiveness Act of 1998, 100-418, 5121 (15 U.S.C. 278k).
- Osborne, D., & Gaebler, T. (1992). *Reinventing government: How the entrepreneurial spirit is transforming the public sector*. Reading, MA: Addison-Wesley.
- Porter, M. (1990). *The competitive advantage of nations*. New York: Free Press.
- Scheirer, M. A. (2000). Getting more "bang" for your performance measures buck. *American Journal of Evaluation*, 21(2), 139-149.
- U.S. Census Bureau. (2000). *1998 annual survey of manufacturers: Statistics for industry groups and industries*. Suitland, MD: Author.
- Voytek, K. P. (2001, October). *Measuring the effectiveness of economic development programs: The Manufacturing Extension Partnership's effect on the productivity of small manufacturers*. Paper presented at George Washington University's Institute of Public Policy, Washington, DC.
- Wholey, J. (1977). Evaluability assessment. In L. Rutman (Ed.), *Evaluation research methods: A basic guide* (pp. 41-56). Beverly Hills, CA: Sage.