Business Services for Small Enterprises in Asia: Developing Markets and Measuring Performance

International Conference
Hanoi, Vietnam - April 3-6, 2000

Evaluation as an Effective Management Tool

By
Eric Oldsman

Nexus Associates, Inc. with support provided by the World Bank

Sponsors
German Agency for Technical Cooperation (GTZ)
United States Agency for International Development (USAID)
Mekong Project Development Facility (MPDF)
Committee of Donor Agencies for Small Enterprise Development

Evaluation as an Effective Management Tool

Eric Oldsman, Ph.D
President, Nexus Associates, Inc.

Abstract

Donor organizations and other stakeholders are interested in learning more about the performance of business development services (BDS) that have been established to help meet the diverse needs of small and medium-sized enterprises (SMEs) in different countries. To this end, the Committee of Donor Agencies for Small Enterprise Development has developed a preliminary Performance Measurement Framework (PMF) to guide research on BDS performance. The PMF presents a set of measures to assess the performance of BDS programs with the eventual aim of establishing performance standards for programs around the world.

This paper provides a critique of the PMF within a broader discussion of the role of evaluation, particularly in terms of its use as a tool that managers can use to improve the performance of their own BDS organizations. It suggests that the PMF can be useful in helping program managers and other stakeholders assess performance on various dimensions. However, it also urges the donor committee to exercise considerable caution in establishing common performance standards given substantial differences among programs as well as trade-offs inherent in program designs. It recommends that greater emphasis be placed on in-depth program evaluations grounded in explicit theories of particular initiatives.

Introduction

Nations face enormous challenges in raising the living standards of all citizens. Many of these problems relate to the ability of the economy to generate gainful employment along with rising productivity. In response, numerous programs have been launched throughout the world over the past decade with the intention of improving the performance of small and medium-sized enterprises (SMEs).[1] Since SMEs account for a significant percentage of value added and employment in all countries, the assumption is that improved performance in this sector of the economy is critical to future economic stability and growth.

Donor organizations and other stakeholders are interested in learning more about the performance of business development services (BDS) that have been established to help meet the diverse needs of SMEs in different countries. This concern stems from both pressure to ensure that monies are well spent and a desire to strengthen donor-supported programs over time. In light of these concerns, the Committee of Donor Agencies for Small Enterprise Development has developed a preliminary Performance Measurement Framework (PMF) to

[1] In Mexico alone, a recent study identified more than 200 federally supported programs oriented toward SMEs.

guide research on BDS performance.[2] The PMF presents a set of measures to assess the performance of BDS programs with the eventual aim of establishing performance standards for programs around the world.

This paper provides a critique of the PMF within a broader discussion of the role of evaluation, particularly in terms of its use as a tool that managers can use to improve the performance of their own BDS organizations. It suggests that the PMF can be useful in helping program managers and other stakeholders assess performance on various dimensions. However, it also urges the donor committee to exercise considerable caution in establishing common performance standards given substantial differences among programs as well as trade-offs inherent in program designs. It recommends that greater emphasis be placed on in-depth program evaluations.

The review is based in part on work undertaken by Nexus Associates, with support provided by the World Bank, to develop a research, planning and evaluation system for SME programs in Mexico. It also draws on evaluations of BDS programs in the United States and Latin America that the firm has undertaken to illustrate important methodological issues. At the outset, it is important to recognize that it focuses on organizations that define their *primary* mission as strengthening the competitiveness of SMEs while expanding the overall market for business services. In this way they differ from for-profit companies that may provide business services, but whose primary purpose has much more to do with their *own* profitability than the achievement of broader social goals.[3]

The need for good performance measures and accurate measurements

A tremendous amount has been written about the importance of measuring performance. Pick up any business publication or journal of public administration and one will find numerous articles extolling the virtues of performance measurement systems. The basic premise is that managers in all organizations – for-profit companies, not-for-profit institutions and government agencies – need good information in order to run their organizations effectively on a day-to-day basis and make decisions with respect to future directions and needed changes.

There is a saying that is often cited as a rationale for setting clear objectives: “if you don’t know where you are going, any road will do.” There is a corollary that provides justification for measuring performance: “if you don’t know where you are, you’re lost.” Managers need hard, objective data to figure out where their organizations are and where they are heading.

However, while few would argue against the need for performance measures, there is often considerable debate over *what* should be measured and *how* it should be measured. It is generally difficult to develop an appropriate set of measures for a particular program. In this case, the Donor Committee is looking for a *core* set of measures that are appropriate for *all*

[2] “Measuring the Performance of Business Development Services for Small Enterprises,” Mary McVay, September 1999. The PMF was developed with support provided by the International Labor Organization (ILO) and the United States Agency for International Development (USAID).
[3] That said, many of the issues raised in this paper also apply to evaluations of private for-profit businesses.

BDS programs, regardless of differences in goals, target population, geographic coverage, nature of services, scale of operations, and sources of funding.

To be useful, performance measures need to be relevant to specific questions.

It is important to consider how different performance measures might be put to use. Different stakeholders – including funding agencies and BDS organizations – may be interested in addressing a broad range of questions. Measures can be developed with various aims in mind, including the following:

- Characterize conditions within markets to learn more about their structure and performance,
- Determine specific needs within a target population of firms in order to design new programs,
- Establish whether existing programs are being implemented as intended,
- Find out whether existing programs are achieving their objectives, or
- Compare existing programs to judge the relative merits of different approaches to addressing specific needs.

It appears that the PMF is intended primarily to address the last two issues. It suggests that BDS programs measure performance on three dimensions – BDS market development, BDS suppliers and BDS customers – as summarized in Exhibit 1.

Exhibit 1. Overview of PMF Framework

BDS Market Development	BDS Suppliers	BDS Customers
Indicators related to changes in market size, number of providers and products, prices, and market penetration	Indicators related to efficiency, financial self-sufficiency and overall cost-effectiveness	Indicators related to satisfaction, changes in business practices and changes in business performance

That said, the framework suggests a number of measures to gauge the development of the market for business services irrespective of the impact of particular BDS programs. The BDS Market Development component suggests that organizations measure the total size of the market, the number of providers, prevailing prices and other characteristics of the market. Conversations with people involved in the development of the PMF suggest that the aim of these measures is to help ensure that BDS organizations are familiar with the markets in which they operate. However, this would typically fall under the rubric of market research and should not be confused with an attempt to judge the performance of BDS programs.[4] As such, it may be appropriate to separate these two functions.

Assuming performance measures can be developed and implemented successfully (see below), the PMF would help program managers and other stakeholders answer questions that are relevant to their ongoing operations. These include the following:

- Are services being delivered to the intended companies?
- Are companies satisfied with the services?
- Are companies benefiting from the services?
- Is the program contributing to an expansion in the market for business services?
- Are program resources being used efficiently?
- Is the program financially sustainable?
- Is the cost of the program reasonable in relation to total benefits?[5]

The framework is meant to be broadly applicable, driven by an interest in comparing different programs. It attempts to find common ground among programs as a basis for common measures and definitions. As discussed in more detail below, it is not grounded in conceptual models of any particular programs.

To be useful, performance measures need to be valid.[6]

Valid, measurable indicators need to be developed for each concept included in any measurement system. Valid measures provide an accurate reflection of the underlying concept that is supposed to be measured. Most of the measures included in the PMF appear to be valid; others are questionable.

For example, the PMF suggests that programs examine the degree to which customers return to purchase more services as an indication of customer satisfaction. The number of repeat customers is defined as the difference between the number of service units sold during the program and the number of SMEs served during the same period. If the point is to determine the number of customers that purchased more than one service since the inception of the program, the indicator is invalid as illustrated in Exhibit 2. According to the PMF indicator, the program would have had six repeat customers (9-3). In point of fact, it only had one repeat

[4] It is possible to develop a set of indicators to judge whether organizations have adequate processes in place to obtain and use market information. For example, the Malcolm Baldridge National Quality Award includes several criteria related to the knowledge of customers and markets.

[5] As will be discussed below, the PMF is probably not suited to answering other questions often posed with respect to programs such as whether alternative approaches would yield a higher return on investment.

[6] There are two other basic criteria by which to judge whether measures are good – reliability and practicality. Measures need to be reliable in that they can be consistently applied with the same result. Put another way, reliable measures are not subject to measurement error. In addition, measures have to be practical in terms of implementation – it has to be possible to obtain data necessary to calculate measures within budgetary, time and other constraints. A full discussion of these issues as they relate to the PMF is outside the scope of this paper.

customer.[7] The indicator suggested in the PMF is really a measure of the *number of services that were purchased by repeat customers* as opposed to the number of repeat customers.

Exhibit 2. Illustration of Repeat Customers

	Number of Services Purchased
Customer A	1
Customer B	1
Customer C	7
Total	9

There are other examples where the issue of validity is subtler. For example, the PMF includes "the number of SMEs purchasing services" as an indicator of market demand. However, it explicitly excludes SMEs that receive services free of charge or are "paid to receive" services.[8] By definition it does not include SMEs that are currently *using* services, but not paying for them. While the measure proposed in the PMF is valid within these specific limitations, it would be legitimate to include companies that are using services at no charge in an estimate of *total* market demand for a particular service.

Similarly, the PMF does not distinguish between SMEs that are "at-need" of services and those that demand services. While need refers to a characteristic or condition of a firm that relates to a particular problem, demand refers to the extent to which SMEs are interested in participating in a program. The willingness-to-pay for services is not necessarily a valid indicator of need. To a significant degree, the number of companies that are willing to pay for services represents a lower bound on the extent of need within the target population. It is fairly safe to assume that companies that are willing to pay for services are likely to be at-need. However, many companies may not seek assistance even though they may be "at-need" according to some objective criterion. Furthermore, an unwillingness or inability to pay for services may be evidence of certain market failures – the very factors that provide a rationale for intervention in the market in the first place.

A number of the indicators included in the PMF are based on the concept of operating costs. For example, supplier sustainability is defined as the percentage of operating costs covered by

[7] The PMF further instructs users to divide the number of repeat customers by the total number of SMEs served in the period to determine the percentage of customers that are repeat customers. In the example, cited in the text, this figure would be 200 percent. In actuality, repeat customers can account for no more than 100 percent of all customers.

[8] This measure may reflect an implicit assumption that BDS organizations should charge for services. In this regard, it is important to note that there are many companies that provide free services as part of their business strategy such as internet gateways that offer free email services to registered users. Demand exists – people use the service – even though the price is zero. Similarly, BDS organizations may set prices at zero to spur demand for particular services as part of a conscious strategy to achieve certain objectives.

client fees (specifically fees paid by SMES). Operating costs are supposed to include direct and indirect costs, but exclude "research and demonstration, fundraising, monitoring and evaluation, and other start-up costs."[9] Perhaps with the exception of start-up costs (which are not defined in the PMF), all of these could be considered legitimate ongoing costs of operating a BDS program. For example, BDS organizations may need to continually invest in the development of new services to ensure their sustainability as demand for particular services is exhausted and needs evolve over time.

Another suggested measure demonstrates the importance of the particular perspective from which measures are viewed in judging their validity. For example, the change in value added (sales less the cost of purchased materials) is included as a measure of "customer benefits" from services. However, Exhibit 3 shows that companies may be able to boost value added, while suffering a reduction in profit. Suppose that as a result of the service provided through the program, a company decided to manufacture an important component internally rather than purchasing it from a supplier located in the United States. However, it required the company to hire additional labor and purchase new capital equipment, shrinking the profit available for owner compensation. While the higher value added generated in the local economy may be desirable from a social perspective, from the standpoint of the SME owner, the project was not necessarily beneficial. This doesn't mean that BDS programs should forego measuring changes in value added; the indicator, however, is not necessarily a valid measure of "customer benefits." More importantly, the PMF does not contain any measures directly related to productivity or profitability – two important dimensions of SME performance.

Exhibit 3. Illustration of Company Benefits

	Sales	Purchased Materials	Value Added	Labor	Capital	Profit
Before	1,000	500	500	250	100	150
After	1,000	400	600	300	200	100

In a related vein, the PMF suggests that BDS organizations calculate the return on investment (ROI) in the program, while rightfully highlighting the difficulties associated with assessing the full costs and benefits of programs. It argues for a simplified approach, defining ROI as the financial benefits realized by SMEs participating in the program divided by the total program costs. Financial benefits, in turn, are defined as the "sum of the increase in value added experienced by SME clients." Here again, perspective matters. It is important to make sure that all of the costs and benefits are accounted for from specific perspectives – SMEs clients, BDS organizations and society as a whole. In this case, the PMF mixes perspectives, comparing the

[9] Start-up costs are undefined in the PMF. These would need to be specified to ensure consistent usage across all BDS programs.

benefits to clients with the costs shouldered by the BDS organization. Left out of the equation are the costs of program participation born by clients such as fees for services and other costs that were associated with derived benefits. In addition, the PMF explicitly excludes program effects that may extend beyond the immediate SME clients. For example, the knowledge that individuals gain from participating in training programs may spillover to other members of the community, leading to higher incomes. Similarly, as SME clients grow, they may need to purchase additional inputs giving rise to multiplier effects in the local economy. In fact, many would argue that these types of externalities provide the basic justification for investments in programs. However, not all externalities are positive. For example, newly trained workers may displace workers without needed skills. Or successful SME clients may take business away from other firms in the local market. These costs would also need to be included in a full accounting of the return on investment from the perspective of society as a whole.

The point of these examples is to demonstrate the importance of making sure that the measures adopted by BDS programs reflect what program managers and other stakeholders intend to be measured. The validity of a particular measure depends on the specific question that is being asked. An invalid (or unreliable) measure can lead people to draw misleading conclusions about the program.

To be useful, resulting measurements need to be accurate and credible.

Assuming that measures are valid, the PMF will still only be useful if program managers and other stakeholders *can* and *should* trust results. The implementation of the PMF needs to be based on sound methodologies. The PMF suggests that "any methodology" may be used to calculate performance measures as long as the researcher explains it. However, not all methodologies will yield accurate and credible results.

Much of the data needed to calculate performance measures included in the PMF will need to come from program records. For example, organizations need to maintain complete records of *all* companies that have participated in the program, including their characteristics (e.g., number of employees, ownership, industrial sector and geographical location) and the nature of services provided to them (e.g., type of service, service provider, purchase price, and purchase date). They also need to maintain financial records on costs and revenues. Unfortunately, many BDS organizations do not have systems in place to record, store and report data in an acceptable and consistent manner. If underlying data are incorrect, resulting performance measures will also be erroneous.

The PMF also assumes that a great deal of the requisite data on client performance will be obtained through surveys of a random sample of companies.[10] The admonition to base performance measures on a *random* sample is necessary to ensure that results are representative of all of the firms that may have been served through the program.[11] There are, however, other cautions that should be raised in relation to surveys. For example, questions needed to be well designed to avoid measurement error and potential bias.[12] BDS organizations also need to obtain a high response rate on surveys, particularly if there is some reason to believe that respondents may differ systematically from those that do not respond in ways that may affect their answers to survey questions. Suppose an organization surveyed its clients, and achieved a 20 percent response rate with *all* of the respondents indicating that they were satisfied with the program. Based on this information, the program manager and other stakeholders might conclude that there was no need for changes in the program. However, if all of the non-respondents were actually dissatisfied with the services received, the true satisfaction rate would be 20 percent rather than 100 percent – clearly a situation that would need prompt attention. To be able to compare results, the same survey instrument and protocol for administrations needs to be used by all BDS programs. Even then, care needs to be taken in comparing results from surveys with different rates of response.

A more fundamental issue concerns the methodology used to determine the effects of BDS programs on SME clients and the broader market, isolating program effects from all the other extraneous factors that may affect performance.[13] Evaluations should provide credible evidence that BDS programs actually *caused* results. This requires being able to compare what happened after services were rendered to what would have happened – the counterfactual – in the absence of such assistance.[14] For example, as noted above, the PMF suggests that BDS programs measure the impact of the program on changes in value added within SME clients. The central question is whether value added was higher than it otherwise would have been if these companies had not received services through the BDS program. This concept is illustrated in Exhibit 5. The difference between value added with the program and value added without the program is the measure of the program's impact.

[10] In practice, it may be possible to obtain data from other sources. For example, Nexus Associates recently completed an evaluation of the Industrial Resource Center program in the United States using data drawn from the Longitudinal Research Dataset (LRD) maintained by the Center for Economic Studies, U.S. Census Bureau. The data is based on annual surveys of establishments throughout the United States. Similarly, the Instituto Nacional de Estadística, Geografía e Informática (INEGI) in Mexico conducts a monthly survey of approximately 6,700 manufacturing establishments. While the sample is weighted toward large establishments to ensure it reflects the bulk of economic activity in the nation, smaller firms are also included. The monthly survey includes approximately 177 variables, including employment, payroll, production hours, value added, and total output. This type of data can be used in analysis of program impacts.

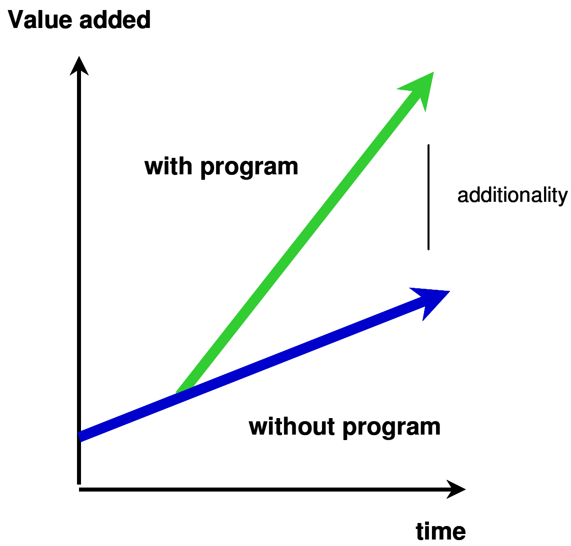
[11] Alternatively, surveys could be sent to all participating companies. Random samples are useful if the number of companies is very high given budget and time constraints. In a random sample each company has a known and independent probability of being included in the sample frame.

[12] The PMF does not include specific questions that should be included in surveys. A common survey instrument would help ensure consistent measures across BDS programs.

[13] This is sometime referred to as attribution or additionality.

[14] In addition, stakeholders may also want to know whether the presumed causal relationship can be generalized to other participants or circumstances.

Exhibit 5. Illustration of Additionality



However, it is impossible to observe the counterfactual directly. As a result, BDS programs need to estimate impacts using various approaches. The strength of causal inferences that can be made about program impacts (and ability to generalize results) depends on the strength of the design. There are three basic approaches to measuring impacts:[15]

- **Before-and-after studies.** This approach entails comparing the performance of SME clients before services are provided and then at another point in time after services have been provided. The difference between the two measurements is taken as the impact of the program. This is the approach that seems to be suggested in the PMF. Unfortunately, despite its obvious appeal as being relatively simple to implement, it may lead to invalid conclusions.[16] It does not take any other factors that may affect

[15] Another approach – case studies – relies on extensive narrative descriptions and other evidence to assert that the program caused observed outcomes. In general, case studies involve multiple sources of information including direct observation, interviews, documents and physical artifacts. In all instances, program logic models play a critical role. Case studies provide rich explanations of how and why the program affected particular firms. However, this approach does not lend itself to simple performance measures.

[16] This approach might be sufficient when there is a clear and close relationship between the program and outcomes of interest. Take the case of COMPITE in Mexico. In this program, certified consultants work with companies over a four day period to improve the performance of one production line along four dimensions – productivity, manufacturing lead time, work-in-process inventory, and floor space requirements. The first day is spent training up to 20 employees in lean production techniques. The second day is spent measuring current performance, diagnosing problems and devising means to improve performance. Day three is devoted to implementing changes. Another set of measurements are taken on the fourth day and compared to those taken before the changes were made. While this is a before-and-after design, because of the proximity of the program intervention there is little doubt that the program caused the changes in observed outcomes. The before-and-after design would not be particularly strong; however, if the question turned to the impact of COMPITE on enterprise-wide outcomes one or two years after the company participated in the program.

performance into account, including industry effects, company maturation or other steps that companies may have taken during the same time period.

- **Participant self-assessment.** The method involves asking SME clients to estimate changes in their behavior or performance due to the program based on their judgment. While this approach attempts to deal with the counterfactual, companies may not be able to estimate impacts correctly and/or may provide answers that are inclined in a certain direction.[17] To be valid, this approach needs to be grounded in a strong program logic model and supported by other evidence.
- **Quasi-experimental design.** The preferred approach is a quasi-experimental design in which the change in the performance of programs participants is compared to similar companies that have not received assistance. A comparison group is constructed and statistical techniques are used to ensure that the two groups are as equivalent as possible on relevant characteristics. To the extent that the two groups are similar, observed differences can be attributed to the program with a high degree of confidence. In general, this approach requires a large sample size and a ready source of comparable data.[18,19]

There are always tradeoffs in selecting an appropriate design to assess program impacts. However, program managers and other stakeholders should be careful placing a great deal of stock in results from weak designs. In all cases, information on program impacts needs to be considered within the context of other information. Here again, care should be taken in comparing results from evaluations that employ different methodologies to estimate program impacts. Differences in estimated program impacts may be due more to differences in methodologies than actual performance differences.

From performance measurement to performance standards

Program managers and other stakeholders are not simply looking for data on the performance of a particular program when they ask questions such as "are companies satisfied with the

[17] The first introduces measurement error making it difficult to reject a null hypothesis even if it is false. The second introduces bias into the estimated value.

[18] This begs the question: how many clients are needed before a quasi-experimental design is feasible. In computing the required sample size, it is important to consider three factors – the likely variance in the outcome measure, the required confidence level, and the desired precision of the estimate. The latter can be thought of as the minimum effect desired by stakeholders – anything less would not be considered successful or of particular interest. A concrete example: assume that stakeholders are interested in determining whether the program has resulted in changes in value added. Further, they believe that the program should aim to increase value added by 5 percentage points more than would have occurred in the absence of the program. Given a standard deviation of 30 percent, a 95 percent confidence interval and a test power of 90 percent, data would need to be obtained from roughly 620 participating and non-participating companies. Assuming a response rate of 50 percent, the program's client base would need to include a minimum of 1,240 companies to use a quasi-experimental design.

[19] See the following recent publications for examples of using quasi-experimental designs in evaluations of BDS programs. *A Record of Achievement: The Economic Impact of the Ben Franklin Partnership*, Nexus Associates, Inc., October 1999 and *The Pennsylvania Industrial Resource Centers: Assessing the Record and Charting the Future*, Nexus Associates, Inc., October 1999.

services?” or “are resources being using efficiently?” They really want to know if the level of performance is in some sense good. This requires an ability to compare the actual performance of a program to some standard such as “at least 90% of clients should be satisfied with services,” or “the cost to provide services should be no more than $10,000 per client.” Establishing a standard means setting a target below which performance will be judged deficient. For example, if the standard for the total operating costs per client were set at $10,000, a program that has higher costs would be considered inefficient.

All programs should establish relevant targets for their own performance as part of strategic and operational plans. This is an important part of management. These standard can be based on past experience, professional judgment and/or comparisons to similar programs. In this regard, the PMF is intended to help organizations establish appropriate standard to govern and judge performance based on comparisons to similar BDS programs around the world.

The issue centers on determining what is an acceptable level of performance for a particular program.[20] The PMF notes that it is likely that performance standards will vary based on differences in program design. Put another way, standards need to be relevant to the particular BDS programs under evaluation. This would entail identifying factors that affect the level of performance relative to the standard, *independent* of the actual operations of the BDS organization. These fixed parameters or control variables define the given environment in which program managers attempt to maximize performance. Examples are shown in Exhibit 6 divided into three classes: market characteristics, target population characteristics and program characteristics. It would be preferable to base the selection of control variables on empirical evidence demonstrating that they actually affect performance.

Exhibit 6. Examples of Factors Affecting Performance

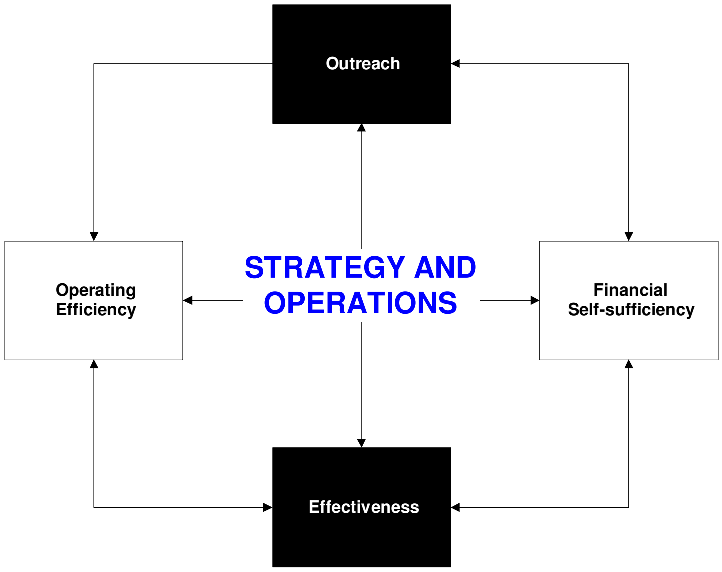
Class	Examples of Control Variables
Characteristics of the market	Size, Buyers, Suppliers
Characteristics of target population	Industry, Firm size, Location
Characteristics of the program	Services, Program maturity, Funding levels

A more important issue is having a clear rationale for establishing particular performance standards. Standards should be tied to the fundamental goals of the BDS programs. In reality, BDS organizations need to balance multiple objectives as shown in Exhibit 7. They need to be

[20] All of the issues raised in the previous section come into play in establishing standards. Performance measures need to be valid and the methodologies used to measure performance need to be sound to avoid the possibility of establishing inappropriate standards. In addition, comparison of the performance of a particular program to the established standard should also be based on the same methodology.

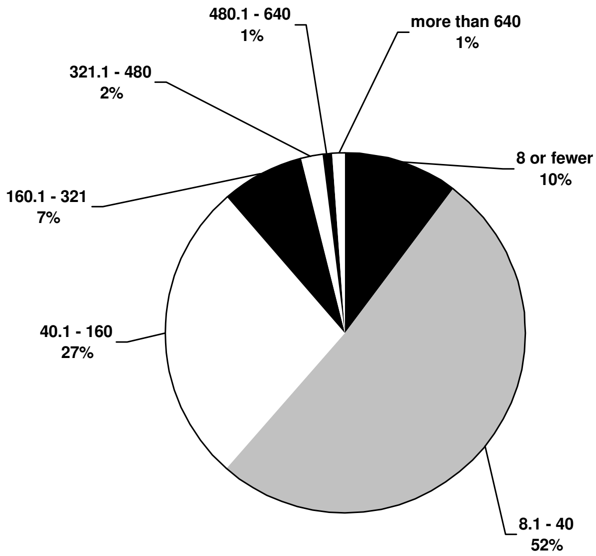
concerned with their financial health, taking steps to make sure that revenues are sufficient to cover their total costs. They also need to ensure that internal processes are efficient, minimizing waste and taking the maximum advantage of available resources. These first two dimensions of performance deal with internal issues, while outreach and effectiveness address the central mission of BDS organizations. Unlike for-profit organizations, BDS organizations have a specific public purpose – reach as many companies in the target population as possible and help improve their competitiveness, while strengthening the overall market for business development services.

Exhibit 7. Multiple Objectives of BDS Organizations



Managers of BDS organizations need to develop strategies and run operations in ways that reconcile objectives within certain constraints. This requires dealing with trade-offs. For example, BDS organizations may take steps to reach a significant number of SMEs by limiting the level of resources committed to each one. A case in point: the Industrial Resource Center (IRC) program was established in 1988 to help small and medium-sized manufacturers upgrade business practices and modernize their production capabilities in order to spur economic growth in Pennsylvania. Since the inception of the program, the IRCs have undertaken approximately 12,000 projects with some 5,000 companies. Roughly 20 percent of clients employ fewer than 10 people; 70 percent have less than 100 employees. At times, IRC staff provide direct assistance based on their own expertise and experience. However, the more typical approach is for staff to serve in a project management capacity, linking manufacturers to qualified third-party service providers and charging for services on a cost-shared basis. As shown in Exhibit 8, almost two-thirds of all client projects required 40 or fewer hours (IRC staff and third-party service providers) to complete.

Exhibit 8. Level of Effort (Hours) Devoted to Client Projects



The emphasis on small projects has enabled the organization to deliver services to numerous small companies. However, almost one-third of all clients report that they did *not* make any changes in their operations as a result of these services nine months after the projects were completed. An analysis of program records and client survey data strongly suggest that the level of effort and mode of service delivery have a direct bearing on the probability of a company reporting that it made changes as a direct result of IRC services.[21] Extensive interactions with the company are essential to success. This suggests that there is a tradeoff between outreach and effectiveness in BDS programs.

Other tradeoffs also exist. For example, the steps that BDS programs may take to generate revenues in order to achieve financial self-sufficiency may run counter to other program objectives. BDS organizations need to establish prices for services – this is a decision that managers need to make. However, at some level prices will prohibit certain companies in the target population from participating in the program or reduce the total investment devoted to particular improvement projects.[22] BDS organization may sacrifice potential impacts in the name of financial self-sufficiency.

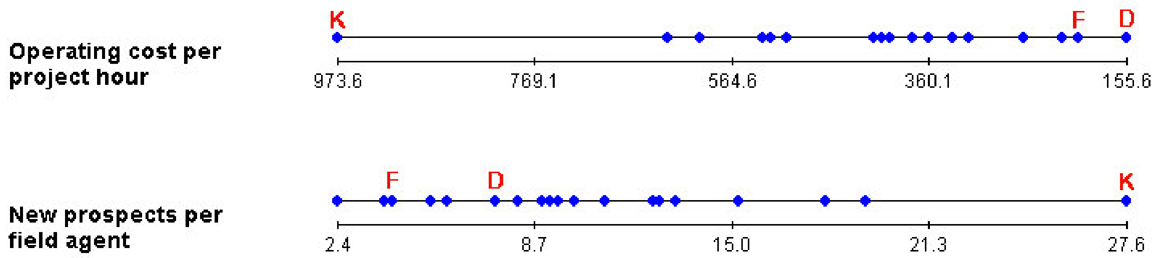
Organizations also need to balance marketing, particularly with respect to reaching out to SMEs that have not yet participated in the program, and ongoing project work. Exhibit 9 suggest that BDS organizations that commit substantial resources to efforts to secure new clients will drive

[21] A probit analysis shows that the coefficient for total project hours and mode of service delivery (one-on-one projects versus group) are positive and statistically significant. In fact, the analysis suggests that 100 hours is needed to boost the marginal probability by four percent, controlling for service delivery mode and employment.

[22] The prevailing market price is the price at which demand and supply are in balance. Prices serve as a way to ration demand.

the total cost of delivering services higher at least in the near term. For example, the operating cost per project hour for Centers D and F are substantially below those of Center K. During the same period, however, Center K was able to generate substantially more prospects per field staff that Centers D and F.[23] The level of performance achieved by each center on each dimension is not inherently better than the others. It depends on the particular objectives of the center.

Exhibit 9. Results of Benchmarking Performance Along Two Dimensions.



Comparisons across different programs can be used to demonstrate that improved performance on certain dimensions is possible. However, the only way to judge whether performance improvements are worthwhile is to consider them within the context of the set of objectives adopted for particular programs. In this vein, development and use of performance standards – the level of performance considered to be good – requires that BDS programs share common objectives and design elements. Given the heterogeneity among programs, this effort is likely to be fraught with difficulties.

From description to prescription

Program managers and other stakeholders are not just interested in determining how well the program is doing relative to a standard. They want to know *how* to improve performance – what changes in the overall strategy or day-to-day operations are needed to boost performance? This is where the PMF falls short. It might be able to indicate areas where performance is deficient; however, it is not capable of pointing to practices that are likely to lead to improvement. In fact, since it is not based on a detailed analysis of actual operations or

[23] Nexus Associates has established a web-based benchmarking system for manufacturing extension centers in the United States. Currently, 40 centers participate voluntarily. Participating centers submit data on a quarterly basis and receive a set of reports showing their performance over time and relative to other participating centers. The anonymity of centers is protected to help ensure data integrity. Projects are defined as work conducted under a signed written agreement between the center and the client. Project hours represent time committed to projects during a specific period regardless of whether this time is actually billed to clients. Operating costs include staff salaries and other operating expenses. In this case, operating costs excludes payments to third parties. A new prospect is a company that the center visited during the period for the first time since the inception of the program. Field staff include people on the center's payroll that have the direct responsibility for working with clients.

conceptual models of individual programs, the PMF cannot be used to distinguish between poor program designs and poor implementation.

Evaluations need to examine operations in great detail.

Details matter. The PMF captures aggregate data, focusing on sums, means and other descriptive statistics. It is not designed to burrow under these figures to get at factors contributing to performance, nor to explore differences in performance within internal processes or among different client groups.

Take a program in Argentina as an example. The Enterprise Development Center in Rafaela-Esperanza began operations in May 1997.[24] From the beginning its operations have been guided by four principles: focus on small businesses, charge for services as a path to self-sufficiency, rely on outside consultants for needed expertise, and forge partnerships with key organizations in the region. The program has been quite successful in reaching companies in the target market and many report significant benefits. Moreover, the organization has reached a point where its client fees are sufficient to cover variable costs associated with client projects, although it still faces a significant operating deficit when fixed costs are included. However, further analysis reveals that the top one percent of its customers as measured by invoiced revenue (roughly six firms), account for 34 percent of the total customer fees generated to date and the top ten percent account for 80 percent of all invoiced revenues. These top customers tend to be larger companies with more than 100 employees that have undertaken one-on-one projects. In practice, this has allowed the organization to cross-subsidize the cost of undertaking projects for smaller firms. This information is essential to judging the organization's performance and to offering insights into how other BDS program might hope to improve their financial situation.

On another front, consider results from an evaluation of the Industrial Resource Center Program in Pennsylvania. An analysis revealed that certain types of companies were more likely to seek assistance than others. Specifically, it demonstrated that the likelihood of a company (i.e., manufacturing plant) becoming a client increased if the company has only a single facility located in a rural area in close proximity to the Center. It also showed that larger plants with growing employment but low productivity growth were also more likely to become an IRC client. This information was important in assessing the actual impact of the program.[25] Perhaps most important, it was useful to understand the factors that contributed to purchasing decisions – information that program managers could use in developing new marketing initiatives to boost participation.

[24] See *Making Business Development Work: Lessons Learned from the Enterprise Development Center in Rafaela Argentina*, Nexus Associates, January 2000. The article was prepared for the Inter-American Development Bank based on an evaluation of the Enterprise Development Center in Rafaela-Esperanza.

[25] The analysis was undertaken to help control for selection bias in an evaluation of the impact of the IRC program on productivity within client firms. The results of the analysis were used in a two-stage least squares model of program impacts.

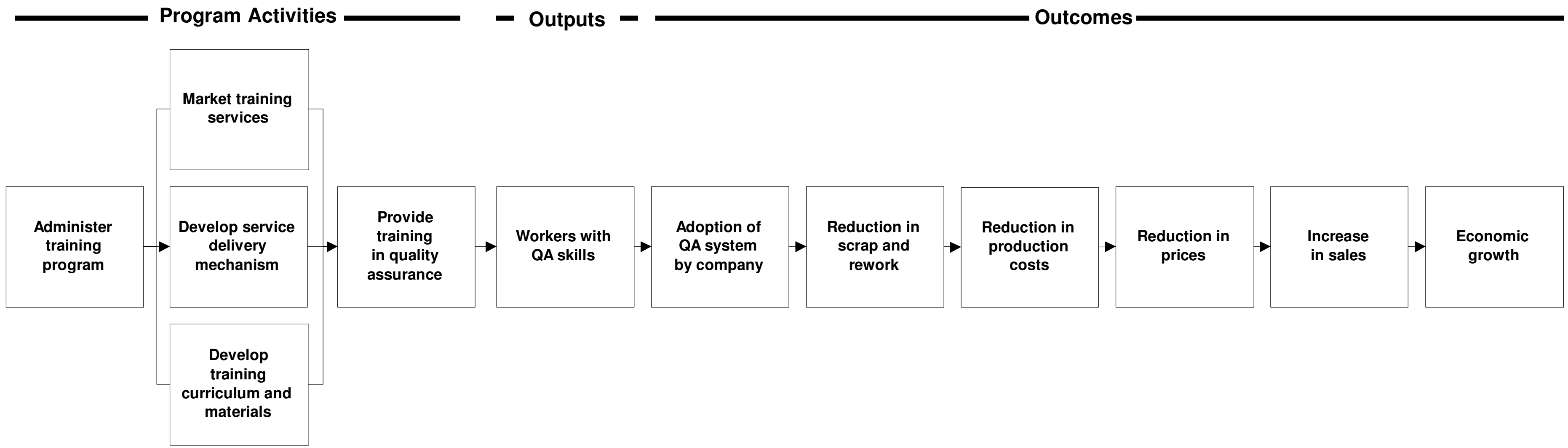
Evaluations should be grounded in theory with detailed descriptions of key components of the program.

While this is the penultimate section of the article, it should really represent the first step in developing a performance measurement system and more comprehensive evaluations. Evaluations should be based on a conceptual model of how programs are supposed to work – what they are expected to do and what they are expected to accomplish. Every program is based on implicit or explicit assumptions regarding the structure, functions, procedures and processes that are necessary to achieve goals. The relationships among program resources, activities and desired outcomes constitute the theory of the program.

As shown in Exhibit 10, program theories delineate the sequence of cause-and-effect relationships that are expected to lead to the attainment of social goals.[26] For example, a BDS program may focus on training workers in quality assurance practices in order to boost sales among SME clients and generate economic growth. In order to deliver the training, BDS organizations need to develop the training curriculum and materials, develop a mechanism for delivering training services and market the training program to companies in the target population. Organizations need to define policies and procedures, allocate resources, and undertake specific tasks to ensure that each process is successful.   These activities are governed by certain assumption and expectations such as “an expenditure of $100,000 will be sufficient to develop high-quality training materials tailored to the needs of companies in the target population,“ “there are three private training firms that are capable of delivering the agreed training program to SME clients,” “as a result of advertisements placed in local newspapers, companies in the target population will become aware of the training program, understand its utility and seek assistance,” or “200 SME clients will send workers for six days of training during work hours at full-pay.” The articulation of program activities provides a solid foundation for developing important questions and performance measures related to program implementation.

[26] The concept of program logic models has been around for a long time. It goes by many names such as program theory, program model, action theory, log frames, etc.

Exhibit 10. Example of Program Logic Model



An articulation of program activities is just the start of the program logic model. Success presumes that the training will actually provide workers with new skills – the principal output of the training program. Furthermore, the model presumes that SMEs will adopt the new quality assurance practice, which in turn will lead to a reduction in scrap or rework, lower production costs, lower prices and higher sales. It is expected that these additional sales will not come at the expense of other firms in the local market, resulting in a net gain in the economy. All of the links in the chain of events are considered necessary and sufficient to yield higher economic growth than would have occurred in the absence of the program. Here again, the definition of the program logic model provides the groundwork for program evaluations. It leads to the questions that should be asked and the specific measures that should be employed in the analysis.

Many BDS programs have not gone through the process of developing fully articulate program logic models, particularly with respect to how the program is intended to achieve social goals. Program managers and other stakeholders often need help in clarifying and refining the model. However, it is imperative that this step be done in close cooperation with other people who are knowledgeable about how the program is expected to work. To improve performance over time, it is important to be able to tell whether the original conception of the program is flawed or whether the organization needs to take steps to strengthen implementation.

Implications for the future

While all programs are launched with good intentions, in reality, the landscape is littered with disappointing attempts to spur business development and economic growth. The PMF is an attempt to provide a structure for assessing the actual performance of BDS programs. Properly executed, it can enable program managers and other stakeholders to base their convictions about different types of initiatives on empirical evidence, rather than preexisting opinions and simple anecdotes.

The development of relevant performance measures and sound measurement systems is critical to good management. With the caveats mentioned above, BDS organizations can use the PMF as a foundation for developing appropriate measures for their *own* programs. In addition, consistent performance measures *among* programs can inform discussions on the strengths and weaknesses of different approaches.

However, considerable caution should be exercised in future attempts to establish and use common performance standards to judge the success of different programs. Programs vary a great deal in terms of their goals and the value that is placed on different objectives. They also differ with respect to basic program elements and operating environments. Performance standards require that programs be fairly homogeneous. Perhaps more importantly, standards should be based on a deep understanding of what really matters in BDS programs. Given the early stage of most programs, it may be premature to assert that certain levels of performance on different dimensions are better than others.

Finally, the PMF is only one tool that can be used by BDS organizations. It is not a substitute for in-depth evaluations that are grounded in the theory of particular programs. In this regard, evaluations should be useful – used by individuals to make decisions concerning broad policy issues, funding allocations, and/or the day-to-day management of programs. Yet most evaluations are conducted to meet the needs of donors for accountability, rather than to provide information that managers need to manage their programs. This needs to change. Evaluations should be oriented foremost to program management.

Program managers are responsible for defining target markets, developing services, delivering services, building partnerships, dealing with personnel, maintaining information systems and handling finances, as well as providing ongoing leadership. In designing evaluations, thoughtful consideration should be given to the nature of the decisions that these people might be contemplating and the information that they would find useful to their deliberations. Evaluations should be relevant to these people, helping them to make better, more informed decisions. This requires detailed analysis of individual programs, influencing how managers think about issues and pointing toward specific actions that managers can take to improve performance. Donor organizations need to consider how they can best help program managers implement needed evaluations and integrate results into planning efforts and ongoing operations. Good intentions will need to be backed by the commitment of real resources.