

Framework for Evaluating the Impact of Small Enterprise Initiatives

Eric Oldsman and Kris Hallberg¹

Abstract

This paper presents a framework for evaluating the impact of donor-funded initiatives designed to promote small enterprise development. We show how to use program logic models to structure the chain of causality between program inputs, outputs, and outcomes. Issues relevant to all evaluations -- attribution, the scope of the evaluation, the choice of indicators and evaluation methodologies, and tradeoffs in evaluation design – are discussed in the context of small enterprise programs. Focusing on interventions to develop markets for business development services, the paper shows how to apply these principles in practice, illustrating different approaches with examples drawn from the literature and previous evaluations of donor-funded programs.

¹ Eric Oldsman is President of Nexus Associates, Inc., an economics and management consulting firm. Kris Hallberg is an economist at the World Bank. Support for the preparation of this paper was provided by the German Agency for Technical Cooperation (GTZ) and the Swiss Agency for Development Cooperation (SDC). The views expressed here are those of the authors and do not represent the policies of GTZ, SDC, or the World Bank Group.

Table of contents

Abstract	i
Table of contents.....	ii
1 Introduction	1
2 Small enterprise initiatives	3
2.1 Rationale and forms of intervention	3
2.2 Targets of intervention and logic models	4
A note on BDS market development	7
3 Outcome measures for small enterprise initiatives	9
3.1 Criteria for selecting measures	9
3.2 Potential outcome measures	10
4 Methods for assessing impacts of small enterprise initiatives	15
4.1 Unit of analysis	15
4.2 Establishing causality	15
4.3 Need for valid comparisons	15
4.4 Four impact assessment methods	17
4.5 Timing of impact assessments.....	24
4.6 Data collection.....	25
4.7 Selection of an appropriate method	26
5 Recommendations and next steps.....	30
5.1 Recommendations.....	30
5.2 Next steps	31
Appendix A – Glossary	33
Appendix B – Selected Bibliography	35

1 Introduction

Donor agencies have invested in a variety of initiatives to promote the development of small enterprises in developing countries.² Some of these programs are designed to expand financing available for small enterprises; others are intended to increase the scope and quality of needed business development services; still others focus on improving the policy environment in which small enterprises operate. The decisions to fund these small enterprise initiatives are based on the belief that well-designed programs will lead to improvements in the performance of small enterprises and associated economic and social benefits.

Initially, many of these initiatives focused on establishing new government institutions and non-governmental organizations to provide various services directly to small firms. However, many donors have become disappointed with these efforts, pointing to inadequate outreach, poor institutional sustainability, and potential distortionary impacts on markets. As a result, in recent years donors have shifted toward a greater focus on providing funding to facilitating organizations to undertake various initiatives to build the capacity of local institutions to deliver financial services and business development services (BDS) to small enterprises, while also stimulating demand for such services among companies in the target population.³ In addition, donors have come to recognize that the formation and growth of small enterprises is dependent on conditions in the business environment in which they operate. As such, efforts have been devoted to encouraging policy, legal and regulatory reforms that affect the ability of small enterprises to compete effectively.

Despite the enthusiasm expressed by program advocates, there has been little effort devoted to rigorous, objective analysis of impacts in terms of desired outcomes. Donor agencies and governments want to know if small enterprises initiatives have actually been successful. In this regard, Donors are under increasing pressure to show that their activities have led to economic growth, employment gains, and poverty alleviation – the fundamental goals of their institutions. Successful initiatives can be expanded and replicated; ineffective programs can be redesigned or eliminated. To aid in this effort, the Committee of Donor Agencies for Small Enterprise Development (Donor Committee) is interested in establishing a framework for evaluating the impact of small enterprise initiatives.⁴

This paper is intended to meet this need. It seeks to provide members of the Donor Committee and other program managers with a better understanding of the critical issues involved in assessing impacts and the various tools that can be used in carrying out such studies. By

² “Small enterprises” are defined to include both microenterprises and small- and medium-scale enterprises. Definitions vary by country, and are usually based on the number of employees, annual turnover, or assets. Typically, microenterprises are defined as firms with up to 10 employees, small enterprises have from 10 to 50-100 employees, and medium enterprises have from 50 to 100-250 employees.

³ Analogous to accepted best practice in the field of microfinance, the BDS market development approach stresses the importance of the commercial orientation and financial sustainability of service providers. Interventions are expected to be short term, with a clear exit strategy defined from the outset. For a description of the BDS market development approach, see Committee of Donor Agencies for Small Enterprise Development (2001).

⁴ The Committee of Donor Agencies for Small Enterprise Development was founded to share knowledge and best practices among donors that support small enterprise programs. Participants include 20 bilateral agencies from 16 countries, 16 multilateral agencies (including the African, Asian, and Inter-American Development Banks), and three international development organizations. We refer only to “donors” as the sponsors of small enterprise interventions, but note that most government-supported small enterprise programs have similar objectives and activities. Thus, the evaluation methods discussed in this paper apply equally to these programs.

definition, the framework focuses on retrospective studies.⁵ Specifically, it focuses on quantitative methods that can be used in summative evaluations of small enterprise initiatives.⁶ In so doing, the paper draws on a rich body of knowledge. The field of evaluation is well established, with professional associations, scholarly journals, conferences, and thousands of published and unpublished studies that have been produced through decades of research. This paper taps this reservoir and applies widely accepted principals and techniques to small enterprise initiatives.

Following this introduction, Section 2 reviews the range of small enterprise initiatives that donors have supported, characterizing interventions in terms of the target of the intervention, goals, instruments, and key cause-effect relationships. This is followed in Section 3 by a discussion of relevant outcome measures for small enterprise initiatives. Section 4 continues the discussion by focusing on different approaches that can be used to assess impacts of small enterprise initiatives, ranging from experimental designs to participant judgment, and offers suggestions on how to decide whether a particular approach is warranted. The paper concludes with recommendations to strengthen future assessment of the impact of small enterprise initiatives and points to areas requiring further work. Two appendices are included. Appendix A contains a glossary of key terms used in the paper. Appendix B contains a selected bibliography of impact assessments and other reference materials.

⁵ In this regard, impact assessments differ from other research efforts designed to assess needs within a target population. Needs assessments typically occur as part of the design process *before* program implementation; impact assessments are carried out *after* program implementation.

⁶ Qualitative assessments are best suited for developing insights into critical process and should be used to supplement quantitative studies.

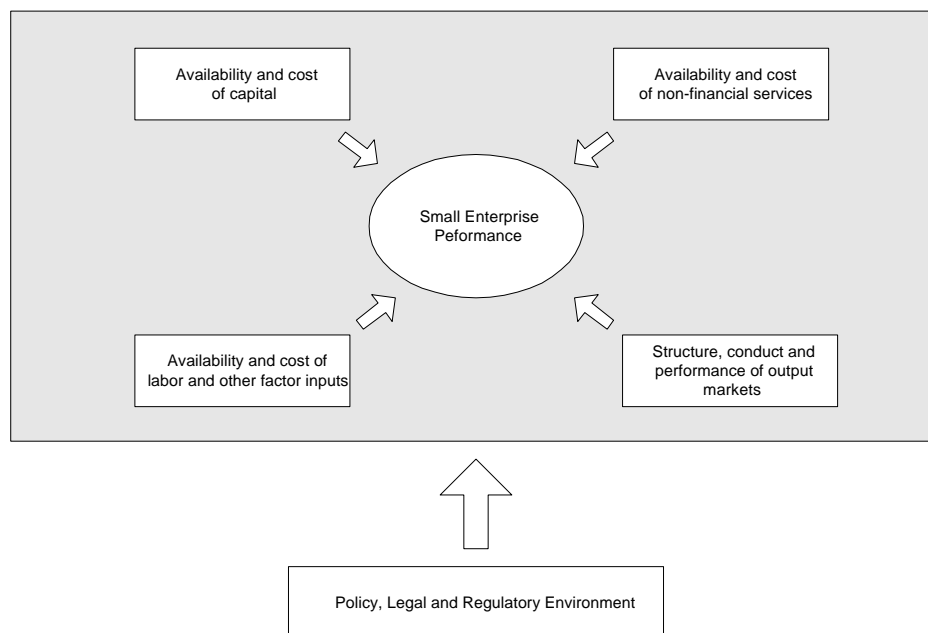
2 Small enterprise initiatives

2.1 Rationale and forms of intervention

Small enterprises play a critical role in the overall economy, providing a broad range of goods and services both for domestic and foreign consumption. In so doing, small enterprises provide an important source of income and jobs in every country. In most developing countries, small enterprises account for the majority of firms and a significant share of total employment. Moreover, many people hold that increased growth among small enterprises will lead to greater demand for labor, rising incomes, and reductions in poverty.

The performance of small enterprises is affected by myriad factors as illustrated in Figure 1. On the one hand, small enterprises need required inputs (e.g., capital, labor, materials, and purchased services) to produce goods and/or services. The availability and cost of these inputs has a direct bearing on the performance of firms. In general, lower factor costs translate into higher profits, all other things being equal. Small enterprises also need to be able to sell their goods and/or services at a price sufficient to earn an adequate return on investment. As such, the performance of firm is dependent on the nature of relevant markets and industries in which they compete. In general, higher profits arise in markets characterized by inelastic demand and limited competition. As noted in the diagram, small enterprises as well as factor inputs and output markets are affected by a broad range of policies, laws and regulations.

Figure 1. Determinants of SE Performance



Given this picture, donors have invested in numerous initiatives over the past decade to address various factors affecting the performance of small enterprises and spur the growth of these firms. As described in Table 1, these initiatives can be grouped under three broad headings – those designed to provide or encourage the provision of financial services to small enterprises; those designed to provide or encourage the provision of non-financial (business development) services to small enterprises; and those that attempt to improve the business environment in which small enterprises operate.

Table 1. Typology of Small Enterprise Initiatives

Focus of intervention	Instruments
Financial services Debt and equity financing Banking services Equipment leasing	Credit lines to financial institutions Consulting and training for financial institutions Direct investment in small enterprises
Business development services Consulting and engineering Management and worker training Marketing assistance Information services	Financial support to BDS facilitators/providers Consulting and training for BDS providers Product development for BDS providers Direct provision of BDS to small enterprises Matching grants/vouchers for small enterprises
Government-mediated business environment Business regulations Property rights and contract enforcement Transparency and corruption Labor policies Trade policies Tax policies	Technical assistance to government agencies Support for public-private dialogue Direct advocacy of specific policies

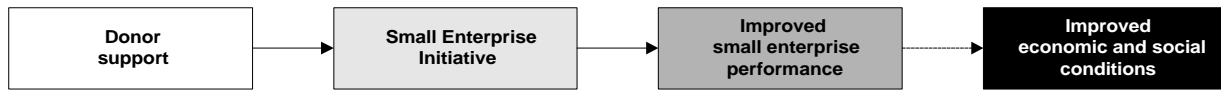
The particular focus and instruments used by donors depends on the overarching mission and policies of the respective organization. Most bilateral aid agencies and non-governmental organizations provide non-reimbursable grants or technical assistance. In so doing, they often work directly with private sector and civil society organizations as well as with government agencies. The multilateral development banks, on the other hand, primarily provide loans to government entities. However, at times, loans are made in concert with non-reimbursable technical assistance supported through trust funds. In addition, the World Bank Group, Inter-American Development Bank and other organizations have units that provide debt and equity financing directly to private enterprises in developing countries.

2.2 Targets of intervention and logic models

Regardless of the particular focus or instruments used, all initiatives aim at boosting the performance of small enterprises and, in turn, improving economic and social conditions as shown in Figure 2. However, it is important to note that while improved economic and social

conditions within a defined region and population are the ultimate goals of donor agencies, small enterprise initiatives are designed to achieve these goals as a result of improved performance among small enterprises.

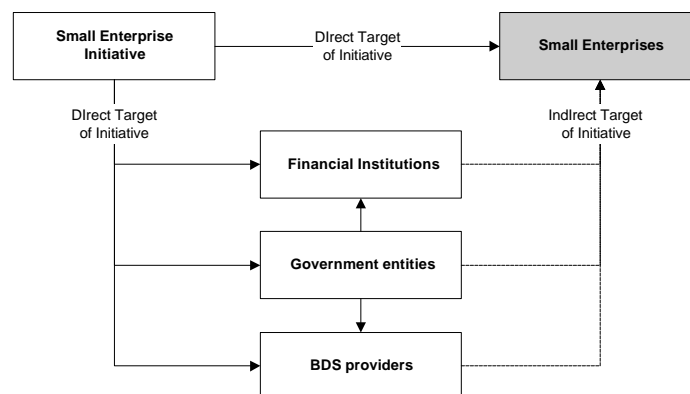
Figure 2. Basic Program Logic Model



As such, small enterprises constitute the principal *target* of small enterprise initiatives. These are generally firms (sole proprietorship or corporation) within a specific size range as defined by employment or assets. Certain initiatives are directed at microenterprises with a handful of workers and virtually no assets; others aim at more substantial companies.⁷ Sometimes, other characteristics or conditions are used to define targeted small enterprises. For instance, particular initiatives may target women-owned firms, rural enterprises, or specific industrial sectors.⁸

However, as reflected in Figure 3, small enterprises may be either *direct* or *indirect* targets of initiatives depending on the particular focus and instruments used. For example, donors may establish credit lines for small enterprises or provide financing directly to particular companies. Similarly, donors may create programs to provide consulting services, training and/or grants directly to small enterprises. In these cases, small enterprises are the immediate subject (direct target) of the intervention.

Figure 3. Direct and Indirect Targets of Small Enterprise Initiatives



⁷ For example, the IFC-managed Mekong Project Development Facility works with firms that have assets and/or planned investments of at least US\$100,000.

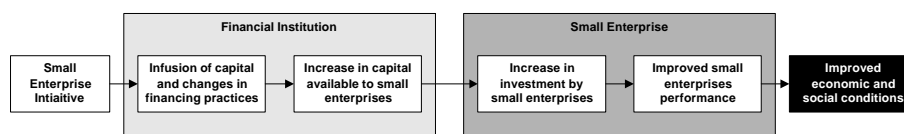
⁸ All too frequently the principal targets of the intervention are ambiguous. For example, the target population of initiatives is often defined in terms of *all* small enterprises in a country. However, in practice, initiatives tend to focus on particular segments of the target population governed by accessibility, relevance, need, and willingness to pay.

In other situations, donor support is intended to change the behavior and/or performance of intermediary institutions that, in turn, have consequences for small enterprises. In fact, a focus on intermediary organizations – financial institutions, government entities and BDS providers – has become the hallmark of the new approach to promoting the development of small enterprises.⁹ As such, the effectiveness of the initiatives is dependent, in part, on their success in transforming these intermediary institutions.

A discussion of the critical causal relationships between activities carried out through the small enterprise initiative and desired outcomes with respect to intermediary institutions, small enterprises, and economic and social conditions is presented below:

- **Financial institutions.** Small enterprises may need access to financing at various stages of their development, including seasonal or revolving lines of credit to meet short-term financing requirements, or intermediate or long-term debt for land, buildings, equipment, and permanent increases in inventory. In response to these needs, donors have invested in efforts to establish and strengthen financial institutions that provide financing to small enterprises. In these cases, the financial institutions are the direct target of donor-supported initiatives. The specific aim of these initiatives is help financial institutions expand available credit, establish new services, improve lending practices, and enhance their overall performance. These efforts are intended, in turn, to result in a greater availability of capital under suitable terms and conditions to small enterprises as shown in Figure 4.

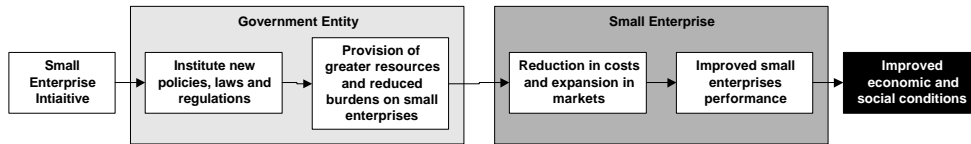
Figure 4. Logic Model for Initiatives Focused on Financial Institutions



- **Government entities.** Small enterprises operate within a policy, legal and regulatory environment established by government (i.e., executive, legislative and judicial branches). In many developing countries, this environment is not conducive to the formation and growth of small firms. As a result, donors have taken steps to work with government agencies to formulate new policies, reduce tax and regulatory burdens, and streamline various procedures in order to provide greater resources and reduce the burdens on small enterprises. As suggested in Figure 5, small enterprises are expected to benefit from these reforms as a result of lower costs (i.e., factor input costs, regulatory compliance costs, and taxes) and an expansion in markets for their goods and services.

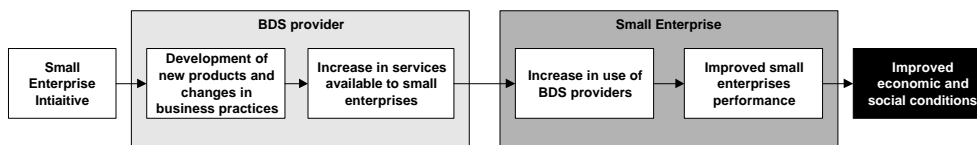
⁹ In addition, some donors have supported efforts to work with policymakers and government agencies to effect changes needed to create an environment more conducive to the growth of small enterprises. These initiatives focus on changing the institutional, legal and regulatory environment in which small enterprises operate. This assistance has generally revolved around the provision of technical assistance and training to government officials. It has also included efforts to promote greater dialogue between policy makers and business groups.

Figure 5. Logic Model for Initiatives Focused on Government Entities



- BDS providers.** Small enterprises may require various forms of outside assistance to address specific needs related to planning, production, operations and other business processes. However, the range of business development services that companies may call on is quite broad, including legal, accounting, auditing, bookkeeping, administrative, data processing, information system design and installation, management consulting, business brokering, market research, testing, marketing, advertising, engineering, and training. (In this regard, services provided to small enterprises are often presented under the general rubric of “business development services.”) As illustrated in Figure 6, donor agencies have taken steps to help existing service providers to develop new services, expand their operations, and strengthen their performance.¹⁰ In these cases, the service providers are the *direct* target of the initiative. It is hoped that efforts to strengthen BDS providers will result, in turn, more services being offered to small enterprises to good result.

Figure 6. Logic Model for Initiatives Focused on BDS Providers



2.3 A note on BDS market development¹¹

Certain donors have called attention to the importance of going beyond strengthening particular BDS providers to promoting the development of sustainable markets for business development service. While these initiatives tend to still target BDS providers and small enterprises, intended outcomes are generally expressed in terms of overall market performance. In this case, donors are interested in inducing a permanent shift in the demand for business services among small enterprise and a commensurate shift in the supply of such services in relevant markets.

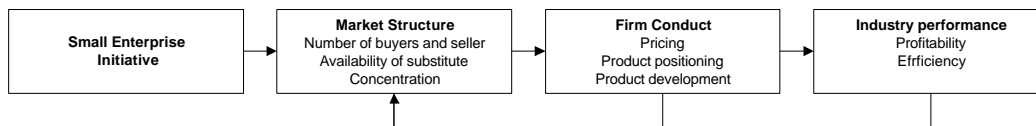
¹⁰ In practice, initiatives tend to emphasize certain services more than others. In this regard, some initiatives are designed explicitly to increase the supply of BDS services tailored to the needs of small enterprises and increase the demand for these services among targeted small enterprises.

¹¹ The discussion of market development applies to financial markets as well.

This begs the question of the definition of the specific market of interest to donors. A market can be defined as trade among buyers and sellers of a *particular* service, in most cases, using money as the common numeraire (as opposed to barter). For example, the market for management training is not the same as the market for industrial engineering services. Moreover, even within the more narrowly defined market for management training, courses may vary significantly in terms of content, method of instruction, quality, duration, time, and location. In a market defined in terms of a particular service, the result of competition is a single price; more than one price suggests an absence of information or other market imperfections.

Markets for particular business development services can be characterized in terms of their structure, conduct, and performance.¹² To a great extent, market structure influences firm conduct which in turn affects industry performance as shown in Figure 7. For example, in a market with few competitors, sellers are likely to increase prices or collude with their rivals. While higher prices may lead to increased profitability, they may also be suggestive of allocative inefficiency as sellers exert market power. However, the casual chain also has important feedback loops. For example, companies may pursue predatory pricing practices in order to drive other firms out of the market, thereby altering its structure. Similarly, supranormal profits in an industry can induce new firms to enter the market – again changing its structure.

Figure 7. Logic Model for Initiatives Focused on Market Development



In a sense, donors efforts to promote the market for business development services is about altering this dynamic to produced desired outcomes. As noted below, these outcomes may relate to the structure of the particular BDS market targeted by the intervention, the conduct of participating companies, and the overall performance of the industry.¹³

¹² See Mason (1939) and Bain (1959).

¹³ From a broader perspective, the intended goal of market development may be expressed in terms of an increase in total surplus (consumer's plus producer's surplus), representing the additional value from an expansion in trade in a particular BDS market.

3 Outcome measures for small enterprise initiatives

3.1 Criteria for selecting measures

The logic models presented above are best thought of as guides. Before embarking on an assessment of the impacts of a particular initiative, however, stakeholders need to specify the particular targets of the initiative and intended outcomes as precisely as possible.¹⁴

With this in hand, the next step is to establish a set of measurable indicators that can be used in assessing the impact of specific small enterprise initiatives.¹⁵ In developing these measures, it is essential to consider the following:

- **Relevance.** Measures selected for the impact assessment need to be germane to the particular initiative being studied.
- **Validity.** Measures need to provide an accurate reflection of the underlying concept that is supposed to be measured.
- **Reliability.** Measures should be subject to as little measurement error as possible.
- **Practicality.** It has to be possible to obtain data needed to calculate measures.

The results of the impact assessment will only be accurate and credible to the extent to which measures are relevant, valid, and reliable. But it also has to be feasible to employ measures given data availability, time, and budgetary constraints. For example, there are a variety of ways to measure productivity – e.g., output per employee, value-added per labor hour, total factor productivity. The last measure reflects the additional value generated through the use of capital, labor, material and other factors of production. While it is arguably the best measure of productivity, it is very difficult to obtain required data even within large companies with sophisticated information systems. On the other hand, although output per employee as a measure of productivity may be misleading given that increased outsourcing will show up as a productivity gain, it is relatively simple to obtain necessary data. On balance this may be the best choice for a specific impact assessment. Like other aspects of designing and implementing evaluation, selecting outcome measures often involves tradeoffs.

In this regard, proxy measures – indirect measures of the outcome of interest – may need to be used in certain circumstances. For example, donors and stakeholders may want to determine whether a particular initiative aimed at boosting quality actually helped small enterprises drive down production costs. While it may be preferable to examine the impact of the program in terms of unit costs, it may not be possible to obtain this data at all or only at enormous cost. Therefore, some substitute needs to be found. In this case, it may be appropriate to use an intermediate measure such as scrap rate or capacity utilization as a proxy for unit costs. Both measures are associated with fixed and/or variable production costs and may serve as a valid surrogate. In a similar vein, a particularly vexing problem in assessing the impact of small enterprise initiatives is the lack of information on profits within private, often family-owned

¹⁴ Impact assessments can also be used to judge whether an initiative produced unintended consequences. For example, a program designed to generate additional manufacturing output may result in additional pollution.

¹⁵ While this paper focuses on performance measures, impacts can also include attitudinal or behavioral changes in targeted enterprises.

businesses. Although not ideal, sales growth may be used as a proxy for profitability over the long-term under the assumption that companies cannot continue to grow if they suffer losses. However, care needs to be taken in the use of proxy measures. For example, customer satisfaction is often used as an indication of impact. This would only be appropriate if satisfaction is highly correlated with sales growth or other intended outcomes. Unfortunately, there is little reason to believe that this is true. Customer satisfaction is likely to be driven more by expectations than actual impacts.

3.2 Potential outcome measures

Outcome measures need to be developed within the context of particular initiatives, reflecting the specific targets and goals of the intervention as well as practical concerns with respect to data availability. *There is no one set of measures that will fit all small enterprise initiatives.*

With that caution in mind, certain measures may be relevant for similar programs. For example, as noted above, all small enterprise initiatives define their principal target as small enterprises. As shown in Table 2, there are a variety of indicators that could be used to assess the impact of particular initiatives on the performance of small enterprises. Many of these indicators focus on changes in quality, turnaround time, and production costs. Others are intended to measure enterprise-wide performance with respect to changes in sales, net profits and employment.

Table 2. Potential Outcome Measures for Targeted Small Enterprises

Indicator	Definition
Defect rate (rework or scrap)	Proportion of units that do not conform to design standards and are subsequently reworked or scrapped
Order-to-delivery time	Total amount of time (hours or days) from receipt of order to delivery at customers' premises
On-time delivery rate	Proportion of orders delivered to customer according to agreed schedule
Customer rejects	Proportion of items delivered to customers and subsequently rejected due to nonconformity
Capacity utilization	Proportion of available resources (e.g., plant and equipment) used in production
Labor productivity	Sales value of output produced during the period divided by direct labor hours used in production
Sales	Revenues derived from the sale of goods or services.
Net profit	Operating profit (sales minus cost of goods sold) and other income less total expenses
Employment	Full- and part-time workers employed by companies or sole proprietorships as of a specific date or pay period, e.g., the week of March 12th

Ultimately, small enterprise initiatives are expected to lead to changes in economic and social conditions with particular regions or countries. As suggested in Table 3, potential indicators center on measures of overall economic activity, the distribution of income, and the extent of poverty within regions targeted by the initiative. In this regard, the rationale for some small enterprise initiatives is sometime presented in terms of “job creation” – generally viewed in terms of employment growth within participating companies. However, the outcome which is ultimately desired is better stated in terms of reductions in unemployment within a specific target population or region. It is possible that employment gains among participants were exactly offset by losses among other companies or were met through an inflow of labor from outside the region. In both situations, the number of unemployed workers in the targeted region would remain unaffected. Therefore, the regional unemployment rate is offered as a potential indicator, supplementing direct employment effects among participating small enterprises.

Table 3. Potential Outcome Measures for Targeted Regions

Indicator	Definition
Output	Sales value of goods and services produced and consumed in the economy
Personal income	Income received by households from employment, self-employment, investments and transfer payments
Income per capita	Personal income divided by the total population in the region
Income distribution	Proportion of total household money income received by households in particular groups, e.g., quintile income groups
Labor force participation rate	Proportion of the working-age population that is employed or actively seeking work.
Unemployment rate	Number of persons without jobs who are available and seeking work divided by the number of persons in the total labor force
Poverty rate	Number of persons or households with incomes below a specified minimum level divided by total population or households
Human Poverty Index (HPI)	The HPI is a composite index introduced by the UNDP that takes three dimensions into account - life expectancy, illiteracy rate, and overall standards of living.

As noted above, some initiatives are directed specifically at improving the performance of financial institutions that serve small enterprises. Table 4 contains measures that are widely used to assess bank performance and which may be appropriate for some small enterprise initiatives.¹⁶ Potential indicators include the value of the outstanding loan portfolio, particularly loans to small enterprises. Depending on the nature of the particular initiative, this could be further segregated into short-term lines of credit and medium- and long-term loans. Other

¹⁶ Many of these indicators have been used in prior studies of microfinance programs.

measures relate to overall loan performance, focusing on the extent to which banks are able to recover principal and interest. The final set of indicators measure overall performance of financial institutions with respect to profit.

Table 4. Potential Outcome Measures for Targeted Financial Institutions

Indicator	Definition
Loans outstanding	Total value of loans (particularly loans to small enterprises or other sub-targets)
Non-accruing loans	Value of loans that have principal and interest payments in arrears
Loss provisions and write-offs	Value of reserves needed to offset anticipated losses due to loan defaults and value of loans cancelled as a result of actual defaults
Collection rate	Amounts due and actually collected during the year divided by amounts contractually due during year
Net profit	Income less expenses
Return on assets	Net profit divided by average total assets
Return on equity	Net profit divided by average shareholders' equity

Certain small enterprise initiative focus on policy, legal and regulatory reforms to create an environment that is more conducive to the formation and growth of small enterprises. Any assessment of the impact of donor-supported initiatives needs to begin with a qualitative assessment of the nature of changes in policies, laws and regulations affecting small enterprises and role of the donor-support initiatives in influencing the changes. Potential measurable indicators of the impact of reforms center on changes in associated costs to small enterprises as shown in Table 5.

Table 5. Potential Outcome Measures for Targeted Government Reforms

Indicator	Definition
Reporting requirements	Number of forms that small enterprises are required to file with government agencies and level of effort required to complete them
Regulatory compliance costs	Costs incurred by small enterprises to comply with government regulations
Business fees	Fees paid by small enterprises for registrations, permits and licenses
Bribes	Bribes paid by small enterprises to government officials
Import tariffs and quota levels	Tariff rate and quotas on imported goods
Tax burden	Effective tax rate for small enterprises after accounting for exemptions, deductions and credits

Small enterprise initiatives may also seek to strengthen the performance of BDS providers in terms of the provision of services to small enterprises and overall sustainability. While there are a host of potential indicators, the core measures of enterprise-wide performance relate to growth in sales, profits and employment as shown in Table 6. Again, given the focus on small enterprises, it would be important to assess the extent to which sales are derived from that segment of the market.

Table 6. Potential Outcome Measures for Targeted BDS Providers¹⁷

Indicator	Definition
Sales	Revenues derived from the sale of services (particularly sales to small enterprises or other sub-targets)
Net profit	Operating profit (sales minus cost of goods sold) and other income less total expenses
Employment	Full- and part-time workers employed by companies or sole proprietorships as of a specific date or pay period, e.g., the week of March 12th

¹⁷ The Performance Measurement Framework also provides a number of possible indicators of BDS provider performance. See McVay, Miehlbradt, and Canedo (2001) and McVay (2001).

Finally, some initiatives may be focused on the development of certain markets for business development services. Given a definition of the specific market, potential indicators with respect to the structure, conduct and performance of markets are provided in Table 7.

Table 7. Potential Outcome Measure of Targeted BDS Market

Indicator	Definition
Number of sellers	Number of companies offering a specific service in the defined market
Number of buyers	Number of small enterprises that purchased a specific service.
Willingness-to-pay	Maximum price as which individual consumers would still want to buy a particular service
Market price	The current price established in the usual and ordinary course of trade between buyers and sellers for the same service.
Volume of services	Quantity of services (value) sold in a specific market. For every sale there is a purchase.
Lerner index	A measure of market power defined as weighted average of each firm's margin with weights given by the firms' market shares
Herfindahl index	A measure of market concentration defined as the sum of market shares (squared) of each firm.
Consumer's surplus	Difference between the maximum prices that consumers are willing to pay and the actual price for all units purchased.
Producer's surplus	Difference between revenue received and the variable costs for each unit sold. Represents a contribution to fixed costs and producer profits.
Total surplus	Sum of consumer's and producer's surplus

Again, the outcome measures presented in the tables above are meant to be illustrative. The decision to adopt particular indicators depends on the precise goals of the small enterprise initiative. Donors and stakeholders need to clarify the specific goals of the initiative with respect to intermediary organizations, small enterprises, and regions *before* developing measurable indicators.

4 Methods for assessing impacts of small enterprise initiatives

4.1 Unit of analysis

Impact assessments are used to determine the effects of the program on specific targets. The targets of small enterprise initiatives can include intermediary organizations, small enterprises, markets, and/or regional economies. These targets constitute the relevant units of analysis for impact assessments, depending on the nature of the particular initiative. As noted above, in some cases the intervention may be directed to more than one type of target. For example, a program designed to boost output in small, women-owned businesses, may also be aimed at reducing unemployment in communities where these businesses are located. In this case, the impact assessment would need to treat both women-owned firms and the regional economy as separate units of analysis. At the outset, however, it is important to note that the central principles and issues involved in assessing the impact of interventions are the same, irrespective of the particular unit of analysis.

4.2 Establishing causality

Impact assessment are undertaken to find out whether a particular initiative actually produced intended outcomes. In this regard, impact assessments are concerned with establishing causality. Did the provision of vouchers *cause* an increase in the demand for business services? Did the low-interest loan program *cause* companies to undertake additional investments? Did the management training course *cause* companies to adopt new business practices? Did regulatory reform *cause* improved performance among companies in the target population? These are the types of questions that decision-makers are asking. Donors fund small enterprise initiatives because they believe that they will result in desired outcomes. Impact assessments are intended to provide credible evidence to demonstrate whether their beliefs were justified and initiatives warrant continued support.

In demonstrating that a particular intervention resulted in a specific outcome, certain conditions need to be met.¹⁸ First, changes engendered through the intervention have to be shown to produce the effect – put another way, the outcome must be responsive to the intervention. Second, plausible alternative explanations for the observed outcome have to be ruled out – rival hypotheses must be disproved. Third, the mechanism by which the outcome was produced has to be explained – a theory linking the intervention to the outcome must be articulated. Finally, it must be possible to replicate the results in similar settings. With proper research, apparent correlations can be translated into credible causal explanations.

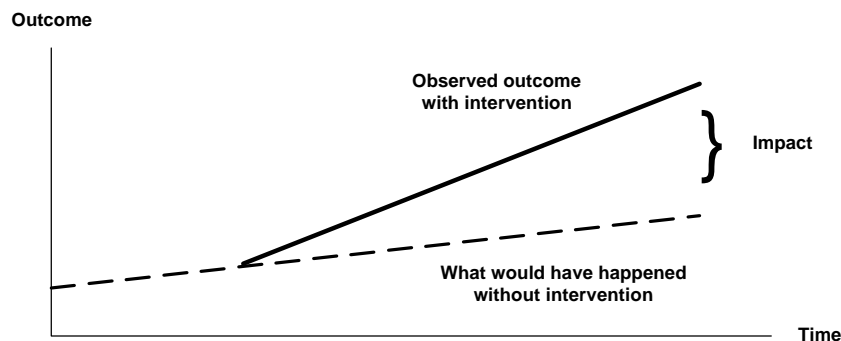
4.3 Need for valid comparisons

The fundamental tenet of impact assessment is the need to compare the observed situation with the intervention to *what would have been* had there been no intervention at all. The difference in resulting outcomes between these two states constitutes the impact of the intervention as

¹⁸ See Mosteller and Tukey (1977) for a discussion of conditions required to demonstrate causality.

illustrated in Figure 8.¹⁹ While the counterfactual cannot be observed or known with complete certainty, the concept of comparing observed outcomes to this hypothetical state underlies all valid approaches to assessing impacts. Valid comparisons imply that the net effect of interventions is isolated from all other extraneous or confounding factors that influence defined outcomes. For example, efforts to improve the performance of small firms by providing vouchers for consulting services may have been undertaken during a time of rapid economic expansion buoyed by substantial tax breaks, aggressive regulatory reform, and booming consumer demand. Given these conditions, it is likely that participating firms would have enjoyed significant growth even in the absence of the voucher program – as a result, the central question is not whether participating firms grew, but rather did these same firms grow more than would have been expected if they had elected not to participate in the voucher program.

Figure 8. The Impact of an Intervention



Thus, the major challenge in impact assessments is to estimate the effect of programs after netting out extraneous factors that affect outcomes. These factors may include specific events or long-term trends in particular industries, regions or countries as in the example cited above. They may also include ongoing developments within participating small enterprises. For example, a program to boost productivity within small enterprises through the provision of technical assistance should take into account the productivity gains resulting from capital investments made by these companies unrelated to the small enterprise initiative.²⁰

Similarly, impact assessment need to account for the voluntary nature of programs. Small enterprises take part in programs of their own volition. Some members of the target population may be more inclined to participate due to greater interest, motivation or other conditions within the firm. This self-selection process can bias results if the factors that lead companies to participate are related to the specific outcomes under study.²¹ For example, initiatives that focus on providing greater access to long-term financing for the purchase of fixed assets are

¹⁹ This is sometimes referred to as “additionality.”

²⁰ Threats to internal validity are generally grouped under several broad categories such as external events, secular drift, maturation, regression and attrition. Readers interested in exploring these concepts in more detail are referred to Cook and Campbell (1979).

²¹ A similar sort of selection bias can occur when organizations select participants based on certain characteristics – this is referred to in the literature as administrative selection.

likely to attract growing companies with progressive management that recognize potential market opportunities, are willing to assume certain risks in the hope of reaping financial returns, and have sufficient collateral to secure the loan. These same characteristics are likely to be associated with future sales growth. It would be inappropriate to compare this segment of the population of firms to other small enterprises that may be struggling to survive. To do so would run the risk of overestimating the impact of the financial assistance program. As discussed below, care needs to be taken to account for potential selection bias in estimating the impact of small enterprise initiatives.

4.4 Four impact assessment methods

While there are numerous variations, the menu of options available to assess small enterprise initiatives is limited to four basic methods based on the type of controls used to isolate program effects from other confounding factors – experiments with random assignment, quasi-experiments with constructed controls, non-experiments with reflexive controls, and participant judgment and expert opinion.²² The strength of causal inferences that can be drawn from the analysis depends on how well the particular approach used in assessing impacts deals with the threats to validity.

Regardless of the purpose or design of the initiative, all impact assessments need to employ one or more of the following methods:²³

- **Experiments with random assignment.** The gold standard in impact assessment is experimental design with random assignment to treatment and control groups. In this approach, small enterprises in the treatment group receive assistance; those in the control group receive an alternative type of assistance or none at all. The critical element of this design is randomization. Random in this case does not mean haphazard; care needs to be taken to ensure that every company has an equal chance of being selected for either group. Random assignment helps guarantee that the two groups are similar in aggregate – any extraneous factors that influence outcomes are present in both groups. For example, both groups can be assumed to be similar in terms of the proportion of small enterprises that are inherently more receptive to making needed changes in business practices. In a like manner, any fluctuations in market conditions can be assumed to affect both groups equally. As such, the control group serves as the ideal counterfactual. Because of this comparability, claims that observed differences in outcomes between the two groups are the direct result of the program are more difficult to refute.

Evaluations using experimental designs are quite common in the health, social welfare and educational arena to test the efficacy of new approaches. (See Table 8.) However, although this approach is very strong, it has not been used extensively in evaluating the

²² A fifth approach -- structured case studies -- can also be used to examine the impacts of intervention on participants and can be used to supplement quantitative techniques. Unlike the other approaches described in the paper, case studies rely on extensive narrative descriptions and other evidence to assert that the intervention 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. While case studies can provide rich explanations of how and why the program affected particular firms, it is difficult to generalize results beyond the firms studied. This is particularly true for programs with diverse clients and services.

²³ Some of the methods to assess the impact of small enterprise initiatives are quite complex, requiring a background in statistics and econometrics. As such, a detailed explanation of the technical issues involved in carrying each type of study is outside the scope of this paper.

impact of initiatives targeting small enterprises. These are several reasons for this. First, political considerations sometimes make it difficult to assign small enterprises to different groups. Second, it is frequently hard to maintain experimental conditions: although small enterprises may be statistically equivalent at the start of the program, some participants may refuse to participate or may drop out of the program. Moreover, the nature of the services provided to small enterprises may change over time. Finally, evaluations using experimental design tend to be costly and difficult to administer.

Table 8. Examples of Experimental Designs

Argentina workfare-to-work experiment.²⁴ The Proempleo program provided a wage subsidy and specialized training as a means of assisting the transition from workfare to regular work. Participants were located in two adjacent municipalities and were registered in workfare programs (mainly the "Trabajar" program). The evaluation's experimental design randomly assigned workfare participants (958 households) to one of three roughly equal-size groups: a) those that were given a voucher that entitled an employer to receive a sizable wage subsidy, b) those that received voluntary skill training along with the voucher, and c) those that received no services and served as the control group.

The evaluation attempted to measure the direct impact of the experiment on the employment and incomes of those who received the voucher and training. A baseline survey and several follow-up surveys were conducted over 18 months. Double-difference and instrumental-variables methods were used to deal with potential experimental biases, including selective compliance. Compared to the control group, voucher recipients had a significantly higher probability of employment after 18 months, though their current incomes were no higher. The impact was largely confined to women and younger workers.

- **Quasi-experiments with constructed controls.** In situations where experimental design is infeasible or impractical, the next preferred approach is a quasi-experimental design. As in the previous design, the change in the performance of participating small enterprises is compared to other similar small enterprises that have not received assistance. However, in this case, assignment to the two groups is non-random. Rather, a comparison group is constructed after the fact. To the extent that the two groups are similar, observed differences can be attributed to the program with a high degree of confidence. Valid comparisons require that the two groups be similar in terms of their composition with respect to key characteristics, exposure to external events and trends, and propensity for program participation.²⁵

There are several types of designs that fall within this general category. These are discussed below in the order of their ability to deal with confounding factors.²⁶

²⁴ Galasso, Ravallion, and Salvia (2001).

²⁵ The issue of the validity of a comparison group is central to this approach. Ideally, the non-participant group should be similar to the participant group with respect to variables affecting outcome measures, but should not have received business assistance through the government program. "Similar" in this context refers to the distribution of values for these variables, i.e., the mean and range.

²⁶ It is important to note that these approaches are not mutually exclusive; they can be combined under certain circumstances to enhance the analysis.

- **Regression discontinuity.** In this approach, scores on a specific measure are used to assign targets to the intervention and control groups in an explicit and consistent manner. The difference in performance between the two groups after implementation is compared, statistically controlling for the variable used in the selection process. For example, scores with respect to the creditworthiness of small enterprises may be used to qualify firms for participation in a loan assistance program – a case of administrative selection. Assuming that an explicit cut-off point is used to determine eligibility, the net effect of the program can be estimated after adjusting for the original selection variable.
- **Statistically equated controls.** This approach employs statistical techniques to ensure that the intervention and control are as equivalent as possible with respect to outcome-related characteristics. In general, this involves using multivariate regression in which the influence of the program is estimated after controlling for other variables that may affect outcomes. For example, the statistical model used to estimate the effect of a consulting program on firm productivity may include various control variables such as firm size, industry classification, geographical location, ownership, initial capital stock as well as factors influencing selection. Selection is addressed through the use of two-stage regression or other techniques involving instrumental variables.²⁷ In the two-stage approach, an initial equation is used to model the selection process. The result of this analysis (inverse Mills ratio) is then incorporated into a second equation along with other control variables to estimate outcomes. As such, this approach explicitly accounts for potential selection bias.
- **Matched controls.** A somewhat less sophisticated approach involves constructing a comparison group that resembles the treatment group as closely as possible based on characteristics considered important in explaining outcomes. For example, companies may be matched based on the same set of variables described in the previous technique. Performance differences between the two groups post-intervention are calculated without further statistical adjustment. However, finding matches for participants based on all criteria simultaneously can be difficult, e.g., another company of the same size, industry, geographical location, ownership, etc.
- **Generic controls.** The last approach uses measurements of performance for the population from which targets are drawn as a control. For example, annual sales growth among participating enterprises may be compared to industry averages with any resulting difference attributed to the program. However, generic controls may not be capable of ensuring comparability with participants and should be used with caution.

Despite their complexity, quasi-experimental designs have been used extensively in evaluating a broad range of development assistance programs. (See Table 9.)

²⁷ See Heckman (1985). A similar technique known as Propensity Score Matching can also be used to control for selection. See Jalan and Ravallion (forthcoming).

Table 9. Examples of Quasi-experimental Designs

Business license reform in Kenya.²⁸ In late 1998, the Government of Kenya instituted a new single business permit (SBP) to replace the existing multiple license system. The Government gave local authorities the option of adopting the SBP system at the start of 1999 or waiting until the beginning of 2000.

By surveying enterprises in three areas that had adopted the SBP system and similar enterprises in three other jurisdictions that choose to retain the old system, the researchers were able to assess the impact of license reform on firms. In each of the six areas, a stratified sample of 60 enterprises was selected for the survey. All told, 262 small enterprises were included in the sample. Basic descriptive data on the enterprises were collected along with data on the licensing process, licensing fees and compliance costs.

The study concluded that the simplification of business licenses reduced compliance costs, but that these benefits were offset by increased licensing fees.

Bangladesh microfinance programs.²⁹ The microfinance programs of Grameen Bank, the Bangladesh Rural Advancement Committee, and the Bangladesh Rural Development Board provide small loans to poor households who own less than one-half acre of land. The programs have served over 4 million poor clients in Bangladesh and have apparently been quite successful.

An evaluation investigated the impact of the programs on 1,800 households in Bangladesh and compared them to a control group of households in areas without any microcredit financing. The study used surveys from 87 villages chosen randomly from 5 subdistricts that served as controls and 24 subdistricts where the programs were implemented. Twenty households were surveyed per village.

The study demonstrated that simple estimates of the impact of programs can be substantially overstated: correction for selection bias nullifies apparently impressive gains. Most of the perceived gains were driven by differences in who gets the loans: they tend to be wealthier and work more than control groups. Once appropriate techniques are used, there is no impact of borrowing on consumption, and children in program areas actually do worse than children in control areas.

Industrial Resource Center (IRC) program.³⁰ The 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. The IRCs are designed to accomplish this mission through a comprehensive set of activities involving a combination of consulting and training services. Since its inception, the state government has committed roughly \$84 million to the program.

A comprehensive evaluation of the program was conducted in 1999. The evaluation included an assessment of the impact of services on participating companies with respect to growth in productivity and output. To help control for potential selection bias, the analysis employs a two-stage procedure.³¹ The first step involves estimating the probability of companies becoming IRC clients as a function of characteristics of the firm. The second step involves estimating the impact of the IRC program on companies after controlling for factors that affect productivity and output growth as well as potential selection bias. The estimated model is based upon a modified Cobb-Douglas production function which includes plant specific factors.

The analysis is based on panel data for individual manufacturing plants – the Longitudinal Research Dataset (LRD) – maintained by the Center for Economic Studies at the U.S. Bureau of the Census. The

²⁸ Abouda and Bowles (2000).

²⁹ Morduch (1998).

³⁰ Nexus Associates, Inc., (1999).

³¹ See the following publications for additional information on procedures used in this analysis: Maddala (1994), Jarmin (1997), and Heckman (1974 and 1979).

dataset provides detailed plant-level data on shipments, employment, factor costs, industry, and other legal and administrative identifiers. It is compiled from the Census of Manufactures carried out every five years and the Annual Survey of Manufactures. The LRD was used to obtain data for both clients and non-clients in Pennsylvania. Companies included in the IRC administrative database were linked to the LRD using a matching procedure developed by the Center for Economic Studies. The matching process identified 2,839 unique IRC client establishments in the census years based on its permanent plant number (PPN). The comparison group included a similar number of companies.

The analysis demonstrated that the program had a significant impact on IRC clients. For the pre-92 cohort, the difference in the growth rate in output and productivity directly attributed to IRC services is estimated at 1.8 percent and 3.6 percent per year over a ten-year period.

Small Enterprise Finance in Ecuador.³² During the 1980s and early 1990s, the World Bank provided Ecuador with over \$140 million in financing for small-scale enterprises through four loans. The objective was to increase the availability of term financing for micro and small firms, while improving the banking system's ability to lend to this sector. The economic environment prevalent in Ecuador during the project period was highly distorted, with negative real interest rates, regulated wages, and a heavily protected industrial sector. In many respects, the objectives of the first two loans could be seen as an attempt to help the government offset some of the more adverse consequences of the Ecuadorian development strategy.

An impact assessment was undertaken to determine the effect of the projects on the performance of SE beneficiaries as well as participating financial institutions. The projects had lent money through the banking system to over 16,000 small enterprises. The evaluation was based on a survey of more than 200 firms -- 144 beneficiaries and 73 that served as a control group. The firms in the control group were chosen randomly from the members of small-scale industry associations in the regions served by the projects. Performance indicators at the end of the project period -- employment, value added, wages, material costs, and capital per worker, as well as data on credit use -- were compared between the two groups. In addition, the evaluation used structured interviews with participating financial institutions to gauge the impact of the projects on SE lending and default rates.

The study found that the rate of job creation among beneficiary small enterprises was lower than that of the control group. Small firms that received credit were less labor-intensive than non-beneficiary firms, and were less efficient (they used more capital and labor to produce the same level of output). In addition, the high failure rate among the beneficiary firms implied that even the limited benefits of the program diminished sharply over time.

- **Non-experiments with reflexive controls.** It is frequently difficult to obtain data on small enterprises that have not participated in the program.³³ In this case, the only option is to base the impact assessment solely on information available for participants, comparing performance before and after the intervention.³⁴ (See Table 10.) In order for results to be valid, the presumption must be made that no other factors – other than the program – contributed to observed outcomes. Because this is seldom true, however, results from studies based exclusively on reflexive controls are often treated with substantial skepticism.

³² "World Bank Support for Small-Scale Enterprises in Ecuador: An Impact Evaluation." Operations Evaluation Department, The World Bank, June 1998.

³³ This is particularly true in the case of full-coverage programs that affect all companies in the target population. Although rare, this would preclude the development of a control group.

³⁴ This could include single or multiple measures both before and after the intervention.

That said, this approach may be sufficient when there is a clear and close relationship between the program and outcomes of interest. Take the case of COMPITE in Mexico: 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-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-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.

Table 10. Examples of Non-Experimental Designs with Reflexive Controls

Kenya agricultural extension.³⁵ The first and subsequent National Extension Projects in Kenya had the dual objectives of institutional development and delivering extension services to farmers with the goal of raising agricultural productivity. A recent impact evaluation attempted to assess the program's impact on agricultural performance. However, the quantitative evaluation was complicated by the fact that the system was introduced on a national scale, precluding the establishment of a control group. The evaluation methodology, therefore, sought to exploit the available pre-project household agricultural production data for limited before-and-after comparisons using panel data methods. The study drew on pre-existing data sources in Kenya (household surveys, participatory assessments, etc.) to construct a panel dataset of 300 households. Institutional development was assessed using staff surveys, interviews, and Ministry of Agriculture reviews of the extension service. Contingent valuation methods were used to estimate farmers' willingness-to-pay for extension services.

The evaluation results show that institutional development has been limited: after 15 years, the effectiveness of extension services has improved little. The service delivery system continues to be costly and inefficient. Extension activities appear to have had little influence on the evolution of patterns of awareness and adoption of recommendations, which indicates limited potential for impact. The data show a small positive change in agricultural technical efficiency in the agricultural sector, but no significant impact of the supply of extension services on productivity at the farm level could be established by using the data in hand.

Although the study had the usual problems of simple before-and-after comparisons, there were several issues that emerged from the analysis -- problems that needed solutions, such as the need to enhance the targeting of services, tailor advice to local social and economic conditions, and increase focus on institutional reform.

It is often the case that programs are not administered uniformly – that is, the intervention may vary in intensity across members of the target population. For example, while some small enterprises may receive 40 hours of technical assistance under a scheme to provide consulting services on a cost-shared basis, others may receive significantly more or less

³⁵ World Bank (1999).

assistance. The impact of varying levels of intensity (sometimes referred to as the dosage effect) can be examined under all four approaches. Depending on the degree of variation, this can strengthen the causal inferences that can be drawn from the analysis. In effect, non-experimental designs with reflexive controls become quasi-experiments. The use of dosage data is particularly important in the case of full-coverage programs where pure comparison groups are unavailable.

- **Participant judgment and expert opinion.** The final approach relies on people who are familiar with the intervention to make judgments concerning its impact. This can involve program participants or independent experts. In either case, individuals are asked to estimate the extent to which performance was enhanced as a result of the program – in effect, to compare their current performance to what would have happened in the absence of the program. While this approach is quite common, it is fraught with problems. It requires people to be able to determine the *net* effect of the intervention based solely on their own knowledge without reference to explicit comparisons. However, it may be the only option available given data and budget constraints. When used, care should be taken to make sure that people consider the counterfactual in their assessment of impacts. (See Table 11.)

Table 11. Examples of Participant Judgment and Expert Opinion

Regional Enterprise Grants.³⁶ The grant program was initiated in 1988 by the Department of Industry and Trade in the UK. The program has two components: a flat rate grant contributing 15% (up to a maximum of £15,000) toward the purchase of fixed assets, and grants covering 50% of eligible projects costs (up to a maximum of £25,000) to develop new product or processes to the point of commercial production. Firms employing fewer than 25 are eligible for participation.

An evaluation of the program was conducted in 1991. The evaluation was based on the results of surveys of a representative sample of 100 firms drawn from the population of companies that has submitted applications after 31 December 1998 and completed projects. All of the surveys were conducted in-person. The survey asked respondents to indicate the impact of the grants in terms of increased turnover, value added, and employment. Additionality was addressed through “in-depth questioning about what firms would have done if no grant had been available and how they would have financed their projects” in the absence of the grants.

The study suggested that the impact of grants for fixed assets was fairly weak, whereas grants for innovation have “performed a useful role amongst firms with genuine growth potential.”

³⁶ Segal Quince Wicksteed, Ltd. (1991).

Impact of Enterprise Law in Vietnam.³⁷ The Government approved the Enterprise Law to simplify requirements for business registration of firms (became effective in January 2000). The law clarifies a number of procedures in the business registration process, eliminates certain requirements imposed on enterprises, and establishes performance standards for government agencies.

The Mekong Project Development Facility (MPDF) commissioned a survey of firms that registered under the new law between January and September 2000. The mail survey was sent to 2,168 enterprises from a list of 6,529 newly registered enterprises in 16 representative provinces and cities. The effective response rate was roughly 10 percent.

The survey asked respondents to indicate the length (days) of the business registration process, the number of days to receive business license from the date of application, frequency of inspections, and attitudes with respect to the regulatory environment.

The study concluded that businesses had benefited from the simplification of the registration process; however, they still faced a number of difficulties in some areas such as obtaining official seals, registering for tax codes, and dealing with licensing requirements.

4.5 Timing of impact assessments

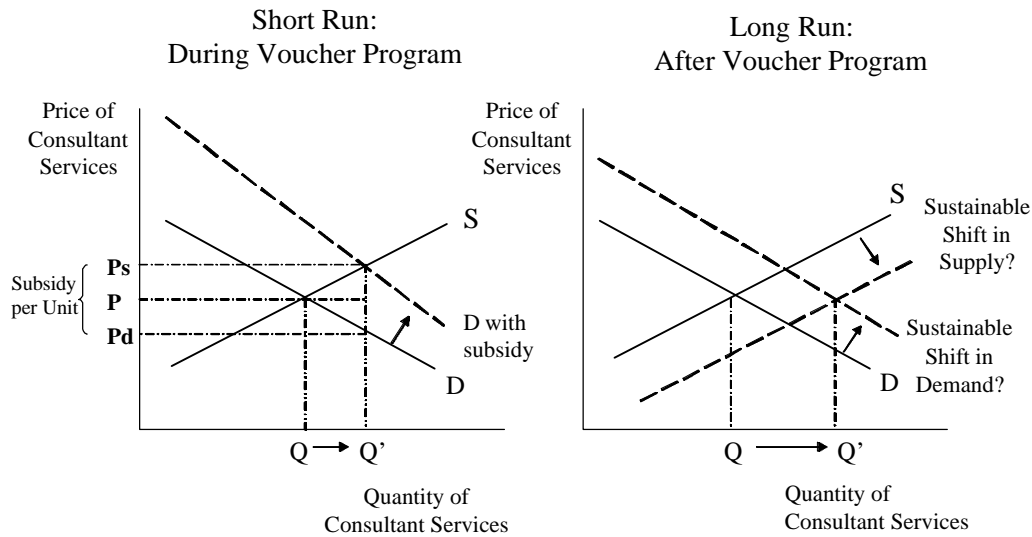
Generally, impact assessments are conducted to satisfy donor requirements. The timing of these studies typically coincides with requests for continued support or at the end of the funding cycle. Given that donor support is commonly provided for three to five years, this usually affords sufficient time for programs to be well-established and most impacts to be realized.³⁸

However, in some cases, it is critical to assess market responses over the long-term. To illustrate, consider the provision of vouchers to small enterprises to offset the cost of consulting services. Donors intend that program will have a significant demonstration effect, leading to an expansion in the market for business services that continues even after the voucher program has ended. As shown in the graph on the left of Figure 9, vouchers can be considered a demand-side subsidy that increases (shifts out) the demand for consultant services. The subsidy lowers the cost of services to small enterprises (from P to P_d) and creates a more attractive price for service providers (from P to P_s). As a result, the quantity of services demanded and supplied in the market expands (from Q to Q'). However, the subsidy-induced expansion in the market may not be permanent. The real question is whether the program results in lasting behavioral change among small enterprises as evidenced by an increase in the willingness to pay for services in the absence of subsidies. The possibility of a sustainable increase in demand is illustrated in the right-hand graph. Sustainable increases in supply are also possible if the larger market for services among small enterprises induces providers of consulting services to innovate. Note, however, that the question of long-term changes in the markets for services – which are fundamental to the BDS market development approach – can only be answered definitively as the voucher subsidy is reduced or eliminated.

³⁷ Mekong Project Development Facility (2001).

³⁸ The time required for effects to materialize depends on the specific outcome of interest. For example, certain intermediate outcomes such as reductions in scrap or rework may show up soon after the intervention. Others such as increased sales may require more time to become visible. At the core, the issue of timing is an empirical question which is amenable to analysis as part of an impact assessment.

Figure 9: Impact of Vouchers on the BDS Market



4.6 Data collection

Regardless of the approach, all impact assessments require accurate and reliable data. At a minimum, it is necessary to collect data on outcome measures for entities that were affected by the initiative. Again, the relevant unit of analysis could be intermediary organizations, small enterprises, or geographical regions depending on the nature of the initiative and particular lines of inquiry. If comparative analyses are contemplated, data on outcome measures and other variables will be needed for members of the treatment and control groups.³⁹ In all cases, baseline data (preferably multiple measures) is needed along with data after the intervention.

There are only two options for obtaining requisite data:

- Draw on existing data maintained by government and other organizations.** Given the cost of data collection, it is preferable to take advantage of existing data to the extent possible. Government statistical agencies in many countries conduct surveys of enterprises and households on a routine basis that might be used in impact assessments. These include national income and expenditure surveys, household income and expenditure surveys, labor market surveys, and various industrial surveys. For example, INEGI conducts a monthly survey of approximately 6,700 manufacturing establishments in Mexico to obtain data on approximately 177 variables, including employment, payroll, production hours, value added, and total output.⁴⁰ Other organizations such as banks, credit unions and cooperatives maintain data on large numbers of companies as part of ongoing

³⁹ Data requirements are specific to the outcome measures of interest and the nature of the analysis that will be conducted. In addition to outcome measures, data may be required for explanatory variables used in regression analyses, including instrumental variables used to control for potential selection bias.

⁴⁰ 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.

operations. These sources should be explored to see whether data required for the impact assessment are available.

- **Commission special surveys.** In many cases, however, the only recourse will be to conduct a survey undertaken specifically for the impact assessment. There are a number of critical issues that need to be addressed to design and administer a survey successfully. The type of survey selected, wording of questions, sampling strategies, follow-up, and data entry procedures all have an important bearing on the accuracy and utility of survey results. A special word on sampling is also in order. It is essential to use probability sampling to ensure valid results; stratification should be considered for greater efficiency and to ensure that the sample accurately represent the overall target population.⁴¹

In addition, all four approaches require administrative records to identify and characterize service recipients as well as the nature of services received. Moreover, certain techniques require additional information. For example, in order to employ regression discontinuity, the program must maintain data on variables used to determine eligibility and/or qualification for participation.

4.7 Selection of an appropriate method

As indicated Table 12, the four approaches to assessing impacts can be applied to a variety of questions that might be posed by donors and other stakeholders with respect to small enterprise initiatives. The table also demonstrates that particular questions can be addressed in a variety of ways subject to data availability, time and budgetary constraints.

For example, the question of whether a particular initiative resulted in a sustainable market for BDS services can be approached in a number of ways. First, it may be possible to provide services such as access to a list of qualified management consultants to randomly selected companies that were interested in obtaining outside assistance. Other companies looking for help in identifying potential consultant would be placed on a waiting list and constitute a control group. Differences in actual expenditures on consultants could be compared to determine the impact of this information. In a related vein, a quasi-experimental approach could be used to determine the extent to which vouchers affected the long-term demand for management consulting services after subsidies were ended. This would involve comparing consultant expenditures of program participants to similar non-participating companies, controlling for other factors such as industry, firm size and geographic location. A third approach would entail comparing the volume of sales of management consulting firms to small enterprises in a particular region before and after the voucher program. Repeated measures pre- and post-intervention could be used to determine the impact of the program. A fourth approach would simply rely on asking managers of small enterprises to indicate whether a particular program made them more or less likely to purchase management consulting services.

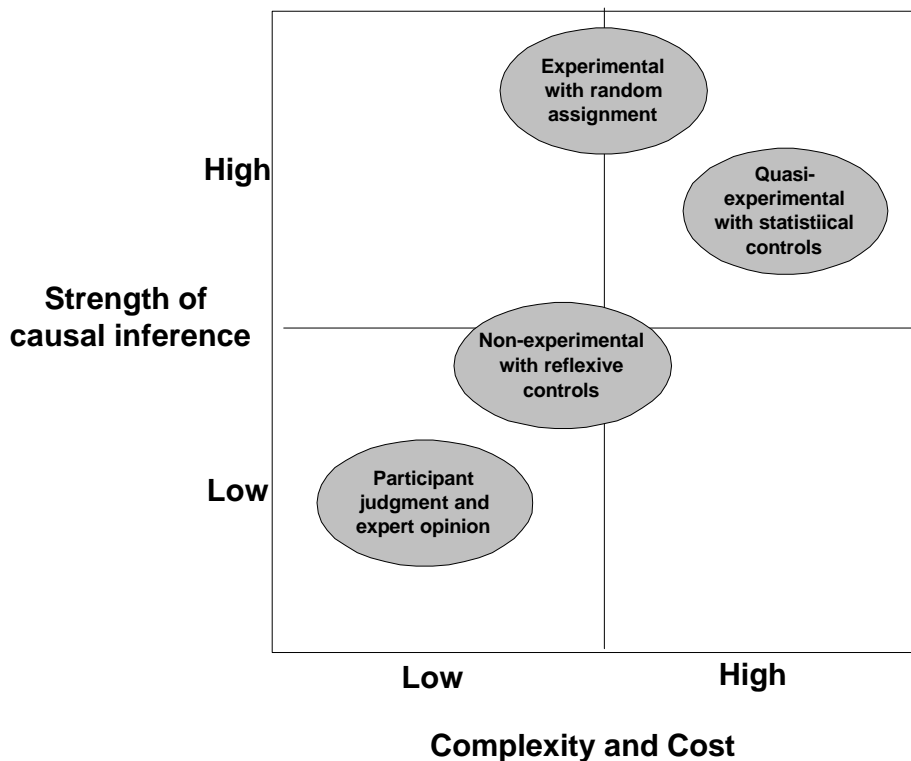
⁴¹ Kish (1965).

Table 12. Alternative Approaches to Assessing Impacts of Small Enterprise Initiatives

Question	Approach	Method
Did the initiative result in a sustainable market for BDS?	Experiment with random assignment	Randomly assign small enterprise to a treatment and control group and compare expenditures on outside service providers over time
	Quasi-experiment with constructed controls	Compare sales growth of service providers in a treatment and control group, statistically controlling for selection and other extraneous factors
	Non-experiments with reflexive control	Compare baseline and post-intervention sales of service providers in the treatment group
	Participant judgment and expert opinion	Ask small enterprises and/or service providers whether they believe that the program has resulted in a sustainable market
Did the initiative result in higher sales among small enterprises that received services?	Experiments with random assignment	Randomly assign small enterprises to a treatment and control group and compare sales growth over time
	Quasi-experiments with constructed controls	Compare sales growth of small enterprises in a treatment and control group, statistically controlling for selection and other extraneous factors
	Non-experiments with reflexive controls	Compare baseline and post-intervention sales of small enterprises in the treatment group
	Participant judgment and expert opinion	Ask small enterprises in the treatment group whether the intervention affected their sales
Did the initiative result in lower unemployment in the region?	Experiments with random assignment	Select comparable regions within the country in which to initiate the program at random and compare unemployment rates over time
	Quasi-experiments with constructed controls	Compare unemployment rates in different regions of the country, statistically controlling for selection and other extraneous factors
	Non-experiments with reflexive controls	Compare baseline and post-intervention unemployment rates in regions in which the program operated
	Participant judgment and expert opinion	Ask experts whether they believe that unemployment rates declined as a result of the program

The choice of the approach to use in a particular impact assessment needs to take several factors into account. Each approach has strengths and weaknesses as illustrated in Figure 10. For example, experimental designs provide strong evidence of causality, but may be expensive and difficult to administer. Non-experimental designs are generally easier to implement, but may not offer strong enough causal inferences.

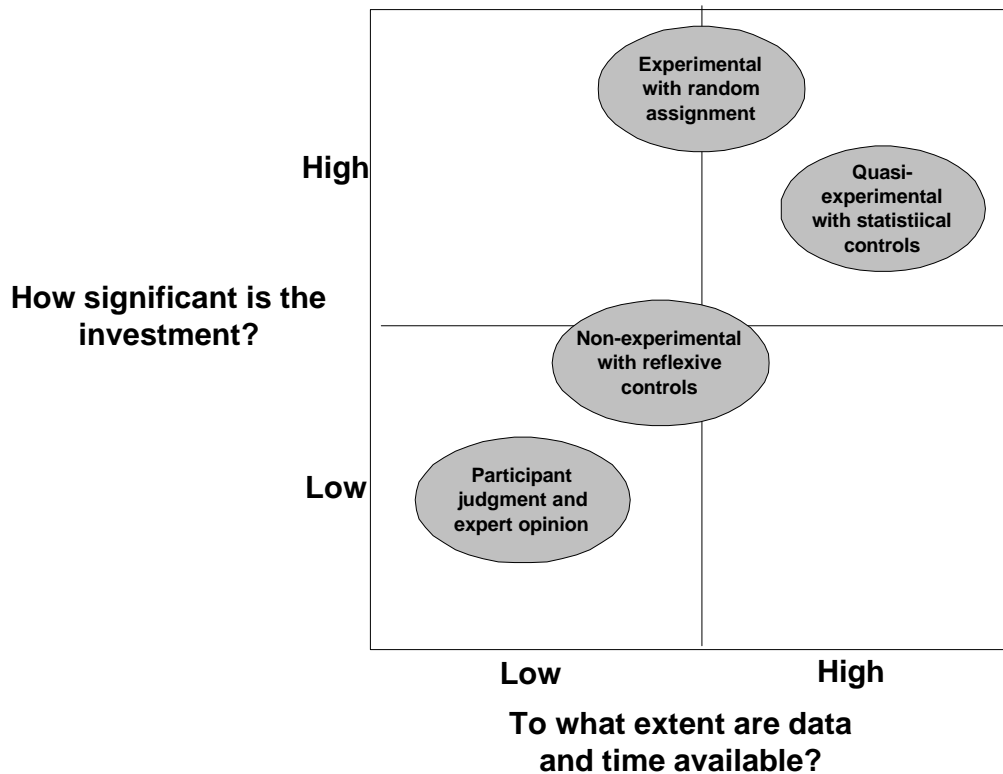
Figure 10. Trade-Offs in Evaluation Design



The figure suggests that more sophisticated approaches such as experimental and quasi-experimental designs should be used wherever warranted given the strength of the causal inferences that can be drawn. However, the additional strength comes at a higher cost. Therefore, these approaches would be appropriate only when further significant investments are being considered. For example, numerous initiatives are established initially as pilot programs with the expressed intention of expanding and/or replicating the initiative if successful. Depending on the magnitude of the required investment, it may make sense for donors to commit resources for a rigorous evaluation to help inform decisions. However, the ability to use these approaches is contingent on whether requisite data can be obtained and whether there is sufficient time to design and implement the study. As such, it is possible to recast the figure

above in terms of two questions as shown in Figure 11.⁴² The general rule should be to use the best possible design from a methodological perspective, taking into account the significance of the investment, as well as practical considerations related to technical feasibility.

Figure 6. Guidelines for Selecting Impact Assessment Method



⁴² Large samples are generally required for quasi-experimental designs. 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 increased sales growth. Further, assume that the program should aim to increase growth by 5 percentage points more than would have occurred in the absence of the program. Put another way, if companies in the comparison group grew by an average of 10 percent per year, stakeholders would like to see participating companies grow by an average of 15 percent annually. 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.

5 Recommendations and next steps

5.1 Recommendations

The basic principals and techniques for conducting impact assessments are well established. However, only in more recent years have they been applied to small enterprise initiatives in developed and developing countries. This experience points to several important recommendations:

- **Clarify targets, goals, and underlying program logic.** Impact assessments require a careful articulation of the targets of the initiative, the specific changes that are expected to be brought about as a result of the initiative, and the causal relationships between particular activities and intended outcomes. This should be summarized in a formal program logic model or log frame. This exercise is best done as part of the process of designing new initiatives, rather than as the initial task of an ex-post evaluation.
- **Plan evaluation at the inception of programs.** Impact assessment should be planned as early as possible, preferably before the initiative has been launched. This is clearly the case for experimental approaches with random assignment; however, in general ex ante designs tend to be stronger since measures for collecting required data can be put in place prior to program implementation. To this end, all programs should be required to develop a formal evaluation plan as a condition for funding. The plan should discuss the purposes of the evaluation; specific questions that will be addressed; evaluation design; process and outcome measures; data collection strategy; possible analyses; reports and other methods of communicating results; timeline; roles and responsibilities of staff and outside contractors, if any; and an estimated budget.⁴³
- **Establish baseline data and program records.** Programs should collect baseline data on characteristics and performance of program targets. Moreover, all programs need to maintain complete and accurate records as part of program implementation, including the nature and magnitude of resources committed to particular companies and/or institutions.
- **Recognize that impact assessment is explicitly about demonstrating causality.** While some people within the donor community are calling for greater accountability in terms of effectiveness, others bemoan the futility of trying to establish attribution. It is difficult to reconcile these views. Impact assessments are concerned specifically with demonstrating that particular initiatives produced the desired results – put another way, they aim to establish causality. All evaluations of small enterprise initiatives should be designed to show effects, rule out alternative explanations, and explain causal mechanisms. The replication of results in similar settings can add further credibility.
- **Build valid comparisons into the analysis.** Assessing the impact of small enterprise initiatives involves comparing observed phenomenon to the counterfactual – a hypothetical situation that would have occurred in the absence of the program. Random assignment,

⁴³ The evaluation plan should go beyond issues related to impact assessments and address other facets of performance, including outreach, operating efficiency, financial self-sufficiency and other process issues.

constructed controls, and/or reflexive controls are needed to isolate the impact of the program from other factors affecting outcomes.

- **Use multiple methods.** The approaches described in the paper are not mutually exclusive. Wherever possible, multiple techniques should be used to assess the impact of particular initiatives. Similar results from different methods can add to the credibility of findings. Moreover, qualitative research should be used to complement quantitative techniques, providing insights into the specific causal mechanisms that come into play in generating outcomes.
- **Recognize that good enough is good enough.** Donors and other stakeholders should strive for as much rigor as possible. However, practical considerations need to be taken into account in designing and implementing an impact assessment. Data, time, and budgetary constraints may make it infeasible to adopt certain approaches. Stakeholders need to accept these limitations.
- **Commit needed resources.** That said, the amount of money devoted to evaluations is at the discretion of donors. Donors must be willing to commit the level of resources needed to design strong evaluations, collect vital data, conduct required analyses, and report results.

5.2 Next steps

Members of the Donor Committee are interested in measuring the impact of their respective small enterprise initiatives in terms of shared institutional goals: economic growth, employment gains, and poverty reduction. However, the task is daunting. Doing a careful job of tracing through the impact of small enterprise initiatives with respect to intended outcomes can be difficult, time-consuming, and expensive. For this reason, members of the Donor Committee may want to consider the following actions:

- **Establish evaluation standards.** The Donor Committee should consider developing a set of standards to guide evaluation efforts in the field of small enterprise development. These standards should be modeled on well-established guidelines that have been adopted by institutions around the world.⁴⁴
- **Collect and disseminate information on good evaluation practices.** The Committee should consider collecting and disseminating information on prior evaluations of small enterprise initiatives. Reviews should focus on the evaluation methodology, describing the specific design, data sources, survey instruments, sampling strategies, and analytical methods, as well as offering a critique of particular strengths and weaknesses. Good evaluation practices should be judged with reference to established evaluation standards. As an initial step, the Donor Committee should fund an inventory and review of existing evaluations to develop a better appreciation of the current state of the art. The inventory could also help identify the largest gaps in knowledge concerning small enterprise initiatives and where it makes sense to focus initial collaborative efforts (see below).
- **Collaborate on comprehensive evaluations of selected initiatives.** Finally, the Donor Committee should consider developing a joint research agenda to answer critical evaluation

⁴⁴ For example, see United States General Accounting Office (1998).

questions concerning small enterprise initiatives. As part of the research agenda, the Committee might consider funding two or three comprehensive evaluations of selected initiatives to determine their impact with respect to intended final outcomes. A few well designed and implemented studies can be used to demonstrate the nature and magnitude of impacts of particular approaches. Subsequent evaluations of *similar* programs might then focus on output and intermediate outcomes, using these major studies as indicative of their likely effectiveness with respect to downstream goals.

For example, many donors support programs that provide technical assistance to BDS providers in order to improve their performance and spur the development of the BDS market. Programs of this type -- with similar target populations, objectives and designs -- could benefit from a single, carefully done impact evaluation of a particular program. Methods for estimating the incidence and distortionary impacts of subsidies, willingness to pay, consumer's and producer's surplus, and "permanent" shifts in supply and demand for BDS could be developed in the process of carrying out the evaluation. Ideally, the evaluation would employ multiple approaches, including experimental or quasi-experimental methods. If the comprehensive impact evaluation suggests that assistance to BDS providers does indeed result in desired social and economic objectives, program evaluations of similar technical assistance programs could focus on outputs and intermediate outcomes such as outreach to target populations, numbers of small enterprise clients, number of service providers, cost per client served, degree of subsidy, etc. These more limited evaluations can be done more cheaply and quickly, giving Donors the means to report frequently on project performance to their funding agencies.

Other types of initiatives that could benefit from this approach include training vouchers for microenterprises; technical assistance to financial institutions to increase small enterprise lending; and policy advice designed to reduce the regulatory burden on small enterprises.

Appendix A – Glossary

- **Confounding factors.** Variables that influence observed outcomes, resulting in an overestimation or underestimation of the true impact of an intervention.
- **Counterfactual.** The conditions that would have occurred in the absence of the program.
- **Control group.** Units (small enterprises, intermediary organizations, or regions) that are similar to those in the treatment group but did not participate in the program.
- **Impact.** The *net* effect of an intervention in terms of specific outcomes after controlling for extraneous confounding factors.
- **Impact assessment.** An evaluation of the extent to which a program resulted in desired outcomes within the target population. Also referred to as impact evaluation or outcome evaluation.
- **Measurement validity.** The extent to which a particular indicator reflects the concept to be measured.
- **Measurement reliability.** The extent to which a particular indicator is subject to measurement error.
- **Outcome.** The results of a program with respect to desired changes within the target population.
- **Probability sample.** A sample drawn from an identified universe where each unit has a known and independent probability of selection.
- **Program logic model.** A description of the causal mechanisms linking program activities to intended outcomes within the target population.
- **Proxy measure.** An indirect measure of an outcome.
- **Selection bias.** A particular type of confounding factor in which preexisting conditions related to outcomes lead certain members of the target population to participate more than others.
- **Small enterprises.** Small enterprises are defined to include both microenterprises and small- and medium-scale enterprises. Definitions vary by country, and are usually based on the number of employees, annual turnover, or assets. Typically, microenterprises are defined as firms with up to 10 employees, small enterprises have from 10 to 50-100 employees, and medium enterprises have from 50 to 100-250 employees.
- **Summative evaluation.** A study undertaken to render a definitive judgment on performance of a particular program particularly with respect to its effectiveness in achieving desired outcomes.

- **Target of intervention.** The unit (e.g., intermediary organization, small enterprise, region) to which a program is directed.
- **Target population.** All units (targets of intervention) in the area served by a program.
- **Treatment group.** Units (e.g., small enterprise, meditating institution, region) that participate in the program.
- **Unit of analysis.** The subject of the impact assessment as defined by the targets of the intervention.

Appendix B – Selected Bibliography

Aboudha, Charles and Roger Bowles (2000), "Business License Reform in Kenya and its Impact on Small Business," *Small Enterprise Development*, Vol. 11, No. 3, September 2000

Bain, Joe S. (1959). *Industrial Organization*. New York: John Wiley & Sons.

Baker, Judy (2000). *Evaluating the Impact of Development Projects on Poverty: A Handbook for Practitioners*. Washington: The World Bank.

Biggs, Tyler (1999). "A Microeconomic Evaluation of the Mauritius Technology Diffusion Scheme." World Bank, Regional Program on Enterprise Development Discussion Paper, November.

Committee of Donor Agencies for Small Enterprise Development (2001). *Business Development Services for Small Enterprises: Guiding Principles for Donor Intervention*

Cook, Thomas D. and Donald T. Campbell (1979). *Quasi-experimentation: Design and Analysis Issues for Field Settings*. Chicago: Rand McNally.

Galasso, Emanuela; Martin Ravallion; and Agustin Salvia (2001). "Assisting the Transition from Workfare to Work: Argentina's Proempleo Experiment." World Bank, September 24.

Heckman, J. (1974). "The Common Structure of Statistical Models of Truncation, Sample Selection, and Limited Dependent Variables and Simple Estimator for such Models," *The Annals of Economic and Social Measurement*, 5: 475-492.

Heckman, J. (1979). "Sample Selection Bias as Specification Errors." *Econometrica*, 47: 153-161

Heckman, James, and Richard Robb. (1985). "Alternative Methods of Evaluating the Impact of Interventions: An Overview." *Journal of Econometrics* 30: 239-67.

Kish, Leslie. (1965). *Survey Sampling*. New York: John Wiley and Sons.

Jalan, Jyotsna and Martin Ravallion (forthcoming). "Estimating the Benefit Incidence of an Antipoverty Program by Propensity Score Matching", *Journal of Business and Economic Statistics*.

Jarmin, R. J. (1997). "Evaluating Impact of Manufacturing Extension on Productivity Growth," Manuscript, Center for Economic Studies, US Bureau of Census, Washington DC. 1997;

Maddala, G.S. (1994). "Limited Dependent and Qualitative variables in Econometrics," *Econometric Society Monographs* No. 3, Cambridge University Press. 1994;

Mason, Edward S. (1939). "Price and Production Policies of Large-scale Enterprises," *American Economic Review* 29: 61-74.

McVay, Mary (2001). "Report on the Outreach and Market Development Indicators Survey." Draft, June.

McVay, Mary; Alexandra Overy Miehlsbradt; and Tim Canedo (2001). "Performance Measurement Framework for Business Development Services: Technical Note on the Research Findings of the Performance Measurement Framework Research." Draft, September.

Mekong Project Development Facility (2001). "Doing Business Under the New Enterprise Law: A Survey of Newly Registered Companies,": Private Sector Discussions, No. 12.

Mohr, Lawrence B. (1995). *Impact Analysis for Program Evaluation*, 2nd ed. Thousand Oaks, Calif.: Sage Publications.

Mosteller, F.M. and J.W. Tukey. (1977). *Data Analysis and Regression*. Reading, MA: Addison-Wesley.

Morduch, Jonathan (1998). "Does Microfinance Really Help the Poor? New Evidence from Flagship Programs in Bangladesh", in Baker (2000).

Nexus Associates, Inc. (1999). "The Pennsylvania Industrial Resource Centers: Assessing the Record and Charting the Future"

Rawlings, Laura (2000). "Assessing Educational Management and Quality in Nicaragua" in Bamberger, Michael, *Integrating Quantitative and Qualitative Methods in Development Research*. Washington, D.C: World Bank.

Rossi, Peter H., Howard E. Freeman, and Mark W. Lipsey (1999). *Evaluation: A Systematic Approach*. 6th Edition. Thousand Oaks, CA: Sage Publications.

Segal Quince Wicksteed, Ltd. (1991). "Evaluation of Regional Enterprise Grants – Second Stage."

Tan, Hong W. (2001a) "Evaluating SME Training Programs: Some Lessons from Mexico's CIMO Program." Washington, D.C.: The World Bank, April.

Tan, Hong W. (2001b). "Malaysia's HRDF: An Evaluation of its Effects on Training and Productivity." Washington, D.C.: The World Bank, April.

U.S. Agency for International Development (2001b). "What Did We Learn from the Pilot Impact Assessment in Kenya's Dairy Products Sub-Sector?" Draft, September 14.

United States General Accounting Office (1998). *The Results Act: An Evaluator's Guide to Assessing Agency Annual Performance Plans*, April 1998.

van de Walle, Dominique (1999). "Assessing the Poverty Impact of Rural Road Projects." Washington: World Bank, processed (summary in Baker (2000))