

Section 4

Evaluation and Impact Assessment

4.1. Planning an evaluation

4.2. Evaluation techniques

4.3. Assessing Impact

4.4. Forthcoming Developments in M&E for BEE

4.5. Key messages

Monitoring and evaluation are complementary and yet distinct aspects of assessing the result of a development intervention. The function of monitoring is largely descriptive and its role is to provide data and evidence that underpins any evaluative judgements. As noted earlier monitoring is ongoing providing information on where a policy, program or project is at any given time (and over time) relative to its respective targets and outcomes. The function and role of evaluation is to build upon monitoring data, bring together additional information and examine whether or not the project results have been achieved.

This section is about evaluation – the *what*, the *who*, the *when* and the *how* questions. It looks at whether BEE reforms have achieved their outcomes (the project ‘purpose’ in logic model terms) and what has been their impact (meeting the project ‘goal’ in a logic model terms). It addresses how to implement good evaluation practices with the use of particular analytical techniques.

It examines strategies and tactics for responding to the challenges of assessing impact, and particularly in answering the difficult questions of:

- What has been achieved and what benefits have come from the changes made by an intervention?
- What, if any, results can be attributed to any given intervention?
- To what extent would changes and results have occurred without the intervention?

4.1 Planning an evaluation

What are the key questions for evaluation?

According to the Development Assistance Committee (DAC) of the OECD³⁷,

*“Evaluation is the systematic and objective assessment of an ongoing or completed project, program or policy, its design, implementation and results. The aim is to determine the **relevance** and fulfilment of objectives, development **efficiency effectiveness, impact** and **sustainability**. An evaluation should provide information that is credible and useful, enabling the incorporation of lesson learned into the decision making process of both recipients and development partners.*³⁸

A comprehensive evaluation therefore typically includes analyzing all five of these criteria. The definitions of these five together with the type of questions asked for each criterion is illustrated in Table 4.1.

Table 4.1: Evaluation Criteria, Definitions and Core Questions

Criteria	Definitions	Core questions	Business registration reform examples
Relevance	The extent to which the aid activity and strategy is responsive to the priorities and policies of the target group, recipient and donor.	<ul style="list-style-type: none"> ▪ Does the intervention address needs? ▪ Is it consistent with the policies and priorities of major stakeholders? ▪ Is it compatible with other efforts? ▪ Does it complement, duplicate or compete? 	<ul style="list-style-type: none"> ▪ Were regulatory issues correctly identified as the key problem for business registration? ▪ Was the type of technical assistance provided appropriate for helping to address the key problems identified?
Effectiveness	The extent to which an aid activity attains its objectives and the degree to which desired outcomes	<ul style="list-style-type: none"> ▪ Are the desired objectives being achieved at outcome and impact/goal level? ▪ Does it add value to what others are 	<ul style="list-style-type: none"> ▪ Has the registration experience improved for enterprises and made it easier to establish a new business? ▪ Has the registration

³⁷ www.oecd.org/dac

³⁸ OECD (2000), Glossary of Key Terms in Evaluation and Results-based Management, pp21

Section 4: Evaluation and Impact Assessment

	are achieved through products and services provided.	doing? <ul style="list-style-type: none"> ▪ To what extent are partners maximizing their comparative advantage? 	experience improved for those institutions implementing these regulations?
Efficiency	The operational and administrative efficiency of projects and services provided.	<ul style="list-style-type: none"> ▪ Are we using the available resources wisely and well? ▪ What is the efficiency of communication mechanisms, knowledge management and coordination with other agencies? ▪ How can we measure outputs – both qualitative and quantitative – in relation to inputs? 	<ul style="list-style-type: none"> ▪ What is the cost benefit analysis of technical assistance inputs against the benefits of: <ul style="list-style-type: none"> - cheaper start-up for enterprise and time saving on the process - lower cost per registration for the implementing registration agency - decrease in time taken to process an application for the agencies
Sustainability	Measuring whether the benefits of an activity are likely to continue after donor funding has been withdrawn.	<ul style="list-style-type: none"> ▪ Will the outcomes and impacts be sustained after external support has ended? ▪ Will activities, outputs, structures and processes established be sustained? 	<ul style="list-style-type: none"> ▪ Have the changes in procedures been 'formally' approved and constituted in the system? ▪ Have the changes in practices been embedded into operations and review systems?
Impact	The positive and negative changes produced by a development intervention, directly or indirectly, intended or unintended.	<ul style="list-style-type: none"> ▪ What changes, positive or negative have occurred? ▪ Are these changes attributable to the initiative? 	<ul style="list-style-type: none"> ▪ Is it cheaper and easier for businesses to start up? ▪ Are more new businesses formally registering? ▪ Does easier business registration mean that new businesses are better performing in early stages? ▪ Has the reputation and client services of the agencies improved?

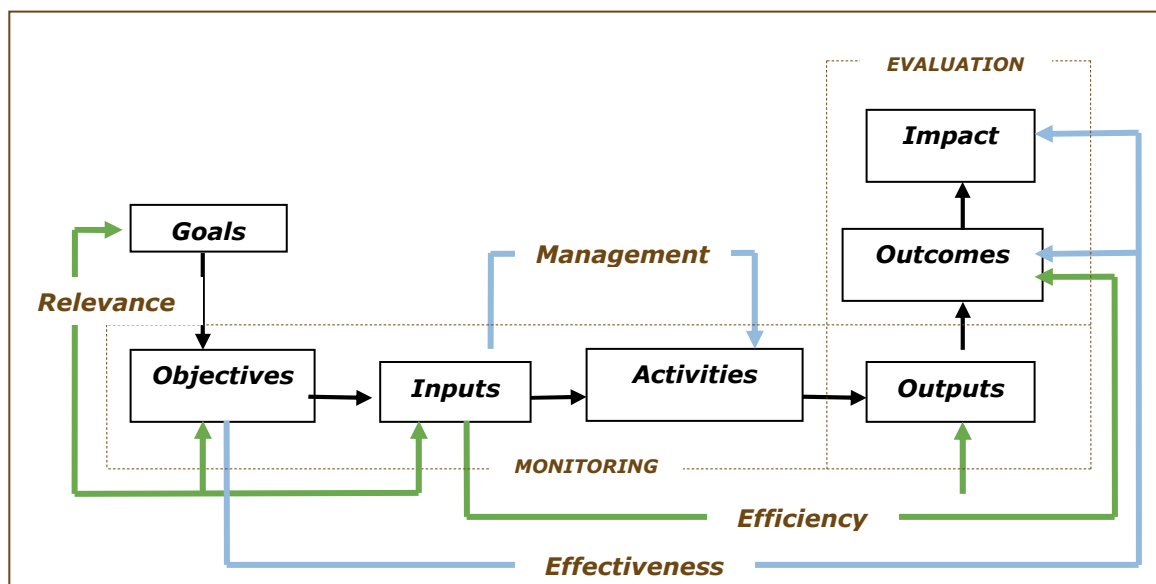
Evaluations can be categorised in several different ways according to when they take place, where they focus and hence what processes they use.

As noted Section 2, the logic model allows for a systematic and diagnostic review of BEE interventions and links M&E indicators and processes to stages of the program cycle

(discussed further in section 5). The core evaluation criteria can also be linked to the LF as shown by Figure 4.1. The intention is to assess:

- The extent of **compliance** and appropriateness of the development partners' BEE objectives and strategy with its overall goals and mandate;
- The **relevance** of the development partners' strategic approach and planned operations for the planned BEE interventions, the management of BEE projects and programs being delivered,
- The **effectiveness** of the BEE activities or the services or technical assistance (TA) provided, and
- The **sustainability** of BEE or investment climate improvements achieved via the services or TA provided.

Fig 4.1: Core Evaluation within the LF and Project cycle



When is evaluation undertaken?

Usually project evaluation is undertaken in line with donor reporting requirements and typically takes place at designated stages in the program cycle (often termed mid-term or project progress review), or immediately after the program intervention is completed (post-program evaluation or completion reporting). Covering all of the core criteria in all evaluations may be an ideal but is not always practical. The evaluation may be conducted at too early a stage to assess impact or sustainability in the longer term. However, in any evaluation it should always be possible to assess some degree of relevance, effectiveness and efficiency as minimum criteria.

The precise protocols and practices of when, what and who is involved in undertaking evaluation and in particular assessing the impact of interventions, varies between development partners and organisations³⁹.

For the purpose of this Handbook the approach for the planning and practice of evaluation is separated into two distinct but interrelated types of activity differentiated by the timing, focus and the methodologies used. They are described as **review evaluations** and **assessing impact** as illustrated in Table 4.2.

Table 4.2: Types of evaluation

<p>Review evaluation</p>	<ul style="list-style-type: none"> ▪ Focuses on outcomes in terms of effectiveness, efficiency and relevance. ▪ Examines whether the activities have delivered the planned outputs and whether these outputs have in turn led to outcomes that are contributing to the purpose of the project. ▪ In DFID for example early reviews are typically called Activity-to-Output Reviews, while later ones are called Output-to-Outcome Reviews. <p>See Annex 4.4 for a sample TOR for a mid term review evaluation</p>
<p>Assessing Impact</p>	<ul style="list-style-type: none"> ▪ Is typically carried out towards or at the end of projects; or after their completion ▪ They usually carried out by those 'outside' of the project in an effort to enhance objective accountability but may also involve insiders in order to enhance lessons learning. ▪ Impact evaluations focus on relevance, effectiveness, efficiency, sustainability in relation to project goals. ▪ Impact evaluations can also be carried out to assess and synthesize the outcomes of several initiatives together on a thematic, sector or program basis to examine their overall impact.

For example, a BEE reform intervention will typically provide various elements of technical assistance to the government in order to achieve specific outcomes (e.g., new enacted legislation leading to an improved investment climate), which in turn would lead to impact (i.e., investment flows, economic growth and employment, and poverty alleviation). The review and impact evaluations looked at different aspects of the 'results achieved' as shown in table 4.3.

³⁹ The key aspects of evaluation practices by IFC, DFID and GTZ are outlined in Annex 4.3

Table 4.3: Review and impact evaluations

Evaluation	Criteria	Measuring
Review	Program Outcomes	Has the policy/regulatory changes been implemented and sustained and the investment climate improved
Impact	Program Goals	Has the better investment climate increased domestic and foreign investment, leading to growth and poverty alleviation

How do we ensure the practice of good quality evaluation?

In general, a good evaluation should aim to meet the generic quality standards as outlined in Table 4.4 which relate to what is involved in evaluation, how it is undertaken, when and by whom. These quality requirements help to ensure that effective and objective assessment practices are undertaken.

Table 4.4: Quality Standards for Evaluation

Standard	Requirement
Utility	The evaluation meets the information needs of the intended users and therefore is relevant and timely
Accuracy	The evaluation uses valid, reliable and relevant information
Independence	The evaluation is impartial, objective, and independent for the process concerned with policy-making, and the delivery and management of development assistance
Credibility	The evaluation is undertaken by evaluators with appropriate skills and experience, is transparent and inclusive
Propriety	The evaluation is conducted legally, ethically and with due regard for the welfare of those involved in the evaluation, as well as those affected by its results
Cost beneficial	The costs of evaluation are proportional to the budget committed to the development intervention being evaluated and remain within the budgetary limits. Resources are used with care

Who should undertake evaluations?

To support these quality criteria, it is important that evaluation activity, especially impact assessment, should be undertaken by those independent of the project or at least those not immediately involved in its implementation. Program officers should be involved in

designing the evaluation as well as contributing inputs to the evaluation exercise alongside other stakeholders, but not undertaking the assessment.

Evaluation teams can include internal officers such as head quarter staff or specialist evaluation staff. For example, IFC have an M&E specialist team in Washington called the 'Results Measurement Unit' as well as regional M&E teams in their Facilities who advise on M&E matters and can be involved in directly evaluating projects. GTZ also has two specialist evaluation units at its head quarters: one focusing on helping project officers to undertake effective M&E of results, the other focusing explicitly on post-project and impact evaluation.

However, evaluations (especially end of project and post-program impact assessment) are activities that are typically undertaken by independent consultants. They bring specialist technical expertise and a sense of objectivity to the evaluation, which are two important criteria for meeting the quality standards noted above. The consultants may come from the private sector or from organizations such as universities research institutes etc. They may be locally based within country or come from internationally operating organizations.

The choice of who undertakes the evaluation of a project and how they are selected and commissioned will depend upon the nature and scale of the BEE reform being assessed. The balance and roles of those internal and external to the project and the practicalities of planning for commissioning and managing evaluation consultants are discussed further in Section 5.

Will who does the evaluation affect diversity and/or inclusion issues?

In Section 1 the importance of ensuring that any evaluation work makes provision for capturing issues of diversity and tries to be as inclusive as possible. Explicit steps need to be undertaken to ensure that this happens throughout the process of designing and implementing the evaluation approach.

Consideration should be given to the questions, which indicators are selected, which target groups are sampled, what research tools are used, who undertakes the research and when and where research takes place. These decisions will all influence the degree to which the diversity of stakeholders will be captured and the level of inclusiveness achieved. Most development partners have practical guidance on these issues, often on intranets.

DFID developed an outline M&E framework, the Integrated Impact Assessment Approach (IIAA)⁴⁰, to provide some guidance on how to ensure that poverty alleviation and other social issues are considered. The approach (see Table 4.5) is based on the logic model. It does not present a new methodology or set of indicators but rather emphasizes three elements of impact assessment⁴¹.

- First, it recommends that impact assessment is brought to the fore in any project/program planning process and that discussions involve consultation with a wide group of stakeholders.
- Secondly, it recommends that any ‘cause and effect relationships’ that are assumed to underpin the proposed BEE intervention are examined and checked with key stakeholders as part of an ex ante proposal. It is at this stage that project designers need to consider impact for a diverse range of groups and in particular how BEE reforms and interventions are likely to impact on the disadvantaged groups. The use of analytical tools such as causal chain analysis and risk assessments should be used alongside participatory evaluation approaches with different stakeholders.
- Thirdly aligned to the above point the IIAA recommends the adoption of a broader ‘lens’ of factors against which impact should be measured. In particular it recommends that consideration is given to social equality and environmental issues alongside the more traditional economic and investment indicators that are held as the primary if not the only success indicators for most BEE reforms.

Table 4.5: The Integrated Impact Assessment Approach

Stage	Tools
Initial screening	<ul style="list-style-type: none"> • Review of current BEE and economic context • Identification of areas to be reformed • Definition of strategy and focus for reform
Program design – ex ante appraisal	<p>Baseline assessment:</p> <ul style="list-style-type: none"> • Review of legislative, policy and regulatory environment • Review of country context and conditions • Consultation procedures and stakeholder analysis • Risk assessment <p>Program design:</p> <ul style="list-style-type: none"> • Determination of policy options that address constraints on the private sector and BEE • Selection of impact indicators – social, economic, institutional,

⁴⁰ Pinder et al (2005) *Guidelines for Assessing the impact of EE programmes: IIAA and Handbook to accompany IIAA guidelines*. for DFID

⁴¹ More information can be found on the website at: www.enterprise-impact.org.uk/BEEenvironments

	<p>environmental</p> <ul style="list-style-type: none"> • Conduct causal chain analysis, assess impact significance • Develop scenarios
Program implementation	<p>Establish monitoring system and ongoing monitoring</p> <ul style="list-style-type: none"> • Focus groups and panels • Point of delivery surveys, score cards • Phone surveys • Mid-term assessment
Program review – ex post evaluation	<p>Output-to-purpose review or purpose -to-goal review</p> <ul style="list-style-type: none"> • Comparison of actual impacts and baseline • Evaluation of implementation and performance • Determine quality of ex-ante assessment

These recommendations and the framework set the agenda for a shift in approach within M&E but it does not prescribe or include a set of core indicators and practices for implementation.

Case snapshot 4.1: Using the IIAA approach

The IIAA approach was applied to the front end design of an IFC business regulatory program in Vietnam. The consultants worked with the local IFC staff, local government officers, businesses and other stakeholders. The 10 day exercise consulting with local stakeholders about the critical regulatory issues and employing an explicit poverty focused approach did not change the fundamentals of the program. However, it did lead to a change of priorities for action, led to set of different results indicators being considered and highlighted a range of important relationships that influenced the degree to which poorer people would benefit from the reforms alongside the business sector.

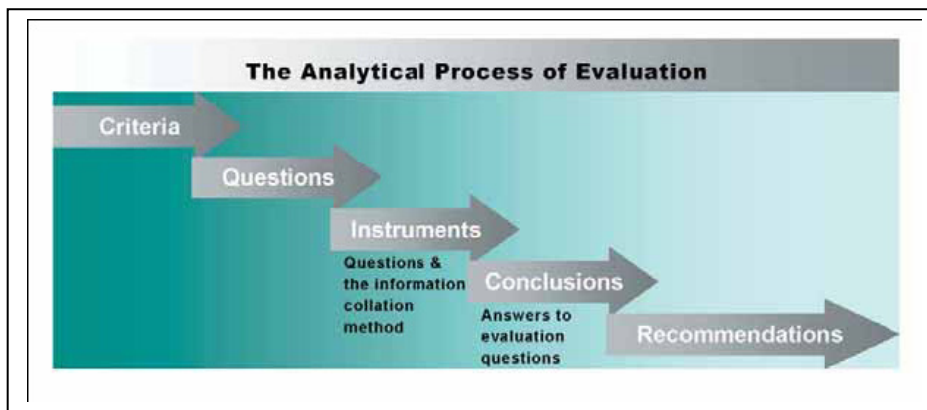
Source: Enterprise Development Impact Assessment Information Service

4.2 Evaluation techniques

What is the starting point?

Undertaking evaluation involves a distinct set of actions requiring specific methods and techniques. DFID in their guidance to officers on project and program evaluation present these as an analytical process of evaluation as shown in Figure 4.2.

Fig 4.2: The evaluation process



The Program officer should:

- Take into consideration the broad **criteria** for development reforms noted earlier (see Section 4)
- Combine these with the **key indicators** identified for the project (Section 2)
- **Identify clear questions** to be addressed by the evaluation
- Make these evaluation questions operational by turning them into **evaluation instruments** for data collection
- Identify the **sources of different data** to be used in the evaluation; and
- Agree the ‘**success rating criteria**’ that will be employed in analyzing the findings from the data collection and the basis on which **conclusions and recommendations are made**.

Which questions should an evaluation prioritize?

An evaluation cannot answer every question that various stakeholders want answered, without becoming burdensome and too time-consuming for those being evaluated and too expensive for those undertaking it. It is important to focus on a set of key questions regarding the output, outcome and impact indicators identified in the Log Frame or plan. These should be set against the core evaluation criteria outlined above.

Case Snapshot 4.2: Developing indicators for business simplification in Egypt

In an IFC PEP MENA (Middle East and North Africa) business simplification reform project in Alexandria in Egypt, a number of output, outcome and impact indicators were developed:

Reform action	Outputs	Outcomes	Impact
Mapping Exercise, Redesign and Implement Processes & procedures	<ul style="list-style-type: none"> ▪ The production of a report with full mapping of existing procedures. ▪ Number of processes mapped. ▪ Number of government authorities engaged in reform efforts. ▪ Number of processes currently under reform. ▪ Investor surveys to assess current business environment. 	<ul style="list-style-type: none"> ▪ Number of laws/regulations changed because of reform work. ▪ Reduced cost and time of registration 	<ul style="list-style-type: none"> ▪ Number of new businesses registered (new investments/ formalized investments) ▪ Additional investment capital generated. ▪ Jobs created ▪ Increase in income ▪ Investor satisfaction with new procedures (based on surveys)

These primary output, outcome and impact indicators can be mapped against key questions for evaluation from the core evaluation criteria:

Criteria	Business registration reform questions
Relevance	<ul style="list-style-type: none"> ▪ Were regulatory issues identified from mapping the key problems faced in the business registration process? ▪ Was the type of technical assistance provided appropriate for helping to address the key problems identified?
Effectiveness	<ul style="list-style-type: none"> ▪ Has the registration experience improved for enterprises? Is it easier to establish a new business? ▪ Has the registration experience improved for those institutions implementing the regulations?
Efficiency	<ul style="list-style-type: none"> ▪ What is the cost of technical assistance inputs against the benefits of: <ul style="list-style-type: none"> - cheaper start-up for enterprises and time saving on the process - lower cost per registration for the implementing registration agency - decrease in time taken to process an application for the agencies
Sustainability	<ul style="list-style-type: none"> ▪ Have the changes in procedures been 'formally' approved and constituted in the system?

	<ul style="list-style-type: none">▪ Have the changes in practices been embedded into operations and review systems?▪ Is there evidence of PPD on these regulatory issues?
Impact	<ul style="list-style-type: none">▪ Is it quicker, cheaper and easier for businesses to start up? And if so:<ul style="list-style-type: none">➤ Can we estimate what are the financial savings by businesses?➤ Are more new businesses formally registering?➤ Is informality decreasing?▪ Does easier business registration mean that new businesses are better established and better performing in their early stages as shown by an increase in capital invested?▪ Have the client services and reputation of the agencies improved?

> For further details, see the full case study on Egypt in Annex 1

What data and information are needed to answer these questions?

Typically evaluation involves using and collecting qualitative and quantitative data sourced from the ongoing monitoring activities of the project, as well as data obtained directly by the evaluation or review team.

Sections 3 looks in detail at the types of secondary data available for the M&E and the key data collection techniques that can be used. Many of the data collection techniques used in evaluation are the same as those that will be used for monitoring, namely: observation, record analysis, interviews and focus groups, questionnaires and surveys. Those more relevant for evaluation are discussed below.

Using secondary data

Key secondary data sources for review evaluation will typically include documentation both **internal** and **external** to the project.

Table 4.6: Documentation sources

<p>Internal project data</p>	<p>Project documentation such as: project design/memoranda and log frame/ impact chain, monitoring/ supervision reports, review reports and documents marking critical incidents or activities in the project implementation.</p> <p>For example: DFID’s guidance⁴² recommends having a audit trail of documentation including evidence of policy changes made or new agreements negotiated by new partner representatives, e.g. after elections. Documents may include key emails as well as more formal letters, reports as well or press cuttings etc.</p>
<p>External data</p>	<p>Reports from partners, other stakeholders, government agencies/departments, research institutes, other development partners, newsletters website notices etc. Statistics from government department and agencies can be critical as background data and providing benchmarks⁴³.</p> <p>For example: Business formalization/informality data. “Most company registration agencies record ‘new registrations’ but fail to record the vast majority of company closures. Some relevant data on closures may be available from the bankruptcy courts, but they tend to exclude the usually larger group of firm that close without going through any bankruptcy procedures. Even if we are only interested in entry, most company registration agencies do not make a clear distinction between a new company and one that is merely changing its name, location, line of activity and/or major shareholders. If we want to track the number of ‘economically active formal companies’ the most reliable and up-to-date source of information in most countries is from the tax authorities.”⁴⁴</p>

Using primary data

In addition to secondary information most evaluations, especially impact evaluations, will involve some form of primary data collection i.e. data specifically collected for the purpose of the evaluation exercise.

Evaluation is usually trying to record the three things:

⁴² DFID (2005): Guidance on Evaluation and Review for DFID Staff Evaluation Department

⁴³ Reference point or standard against which performance or achievements can be assessed

⁴⁴ FIAS (2005): A manual for the identification and removal of administrative barriers to investment, module 6: monitoring, evaluation and continual assessment, pp5-6

- Capturing **quantitative changes in conditions and circumstances** relating to the reform, e.g. the reduction in steps, time and money to register a new business; changes in frequency and duration of business inspections, the frequency and level of fines paid by businesses.
- Capturing **more qualitative changes** in opinions, satisfaction rates, attitudes, e.g. the perceptions of businesses, and of implementing agencies to changes in the regulatory regime
- Capturing **process issues** such as critical incidents and events that have occurred throughout, e.g. the engagement of the business associations in reviewing a reform, the ability of a business association to represent the views of its members, the development of a Public Private Dialogue (PPD) process to improve the quality of regulatory reforms.

Data collection techniques and tools

Not all techniques are suitable for collecting these different types of data as Table 4.7 shows. Data collection techniques must be chosen that are appropriate for the particular research question.

Table 4.7: The strengths and weaknesses of different data collection tools

Method Criteria	Surveys	Rapid appraisal	Participant observation	Case studies	Focus groups
Coverage - scale of applicability	High	Medium	Low	Low	Low
Representative	High	Medium	Low	Low	Low
Ease of quantification	High	Medium	Low	Low	Low
Ability to isolate /measure non-project causes of change	High	Low	Low	Low	Low
Speed of delivery	Low	High	Medium	High	High
Expense of design and delivery	High	Medium	Medium	Low	Medium
Ease of quantification	High	Medium	Medium/ Low	Low	Low
Ability to isolate and measure non-project causes of change	High	Low	Low	Low	Low
Ability to cope with the attribution problem	High	Medium	Low	Low	medium
Ability to capture qualitative info	Medium	High	High	High	High
Ability to capture causal processes	Low	High	High	Medium	Medium
Ability to understand complex processes - e.g. institution building	Minimal	Medium	High	Medium	Medium
Ability to capture diversity of perceptions	Medium	High	Medium	Low	Medium
Ability to elicit views of diverse/disadvantaged groups	Medium	Medium	High if targeted	High if targeted	Medium
Ability to capture unexpected impacts	Low	High	High	High	High
Degree of participation encouraged by method	Medium	High	Medium	Medium	High
Potential to contribute to stakeholder capacity building	Medium	High	Low	Medium to low	High

For example, where the changes in the time, duration and cost of regulative compliance are of interest, then it is valuable to survey a large representative sample of businesses experiencing these regulations. The focus is to capture experiences of compliance in terms of consistent measurable terms such as such as frequency, time and cost. Enterprise surveys (discussed in detail in Section 3) are good for this; case studies are less rigorous as they do not give the coverage of a large number of enterprises.

Case Snapshot 4.3: Using an enterprise survey to measure changes in BEE

The FIAS Regulatory reforms in Latvia used Administrative and Regulatory Cost Survey (ARCS) in late 2001 and again in late 2003 and 2005 to show changes in many aspects of the business environment including for example:

- The frequency and duration of inspections
- The incidence and severity of fines imposed on businesses
- Access to information and updates regarding tax issues
- Changes in the time and cost of administrative procedures, such as the time spent registering a company and registering title transfer in Land Books.
- Number of businesses that regard various specific regulations as an obstacle to the operation and growth of their business

> For further detail, see the full case study on Latvia in Annex 1

Similarly, to assess enterprise satisfaction rates with regulatory issues and services, large-scale surveys are preferable but they tend to be expensive. An alternative way of gaining an insight about these things is to run focus groups (FGs) with key representatives from those groups whose experience the M&E must establish.

Case Snapshot 4.4: Using FGs with business associations in Egypt

The IFC PEP MENA business start up simplification project in Alexandria Egypt engaged the Alexandria Business Association in their reform process throughout the project by using a working group from the association as a sounding board on project progress.

ABA have now decided to undertaken a regular survey of their members to act as a local 'investment climate barometer'. This regular survey from a select but relevant interest group is intended to provide important input to any evaluation work on the project.

> For further details, see the full case study on Egypt in Annex 1

Identifying key stakeholder business associations and groups and inviting representatives along to a FG group will often provide as much insight into business satisfaction with regulations as a survey, especially if the representatives have consulted with their members before they come to the FG.

Case Snapshot 4.5: using FGs with different sector stakeholders in Thailand

The 'Thai-German Programme for Enterprise Competitiveness' (T-G PEC) supports work with businesses in the agro-industry sector in Thailand. The Program utilises a variety of data collection methods using both quantitative tools (such as business surveys, business record keeping) and qualitative approaches (such as focus groups, end of event reviews, meetings) to assess the progress and benefits of their work with enterprises.

Focus groups with enterprises, intermediary business service deliverers and other stakeholders are used to gather feedback on several levels including, enterprise satisfaction levels after specific inputs. They are also used to disseminate results, share knowledge and communicate a range of messages between and magnet different stakeholder groups.

> For further details, see the full case study on Thailand in Annex 1

Can data collection tools be combined?

Evaluation usually involves using a number of different data collection tools to obtain a range of quantitative and qualitative information about the outcomes and impact of a BEE reform. For example, surveys may be complemented by FG discussions and a small number of detailed case studies as well as in-depth interviews with key informants. This performs a checking role or *triangulates* the information collected by combining multiple data sources and methods. In this way, this can help to overcome the bias that comes from only using one source and method of data collection⁴⁵.

Box 4.1: Using triangulation

Triangulation means compensating the use of single data collection methods and a simple study design with the use of several information sources and different methods simultaneously, to generate information about the same topics.

For instance, information from a survey may be supplemented with general experience data from similar interventions, and interviews with a variety of key informants to provide contextual information. In this way the strengths of one methodology can be used to correct or overcome the weaknesses of another and vice versa.

In a situation that affects several parties with different interests, representatives of all parties, as well as some neutral respondents, should be interviewed. This provides a triangulation effect that largely helps to verify information, cuts through conflicting evidence, and reveals insights, in a cost-effective way.

What is a tracer study?

Triangulation is a primary feature of *enterprise tracer studies*. This is where businesses are tracked over a period of time using a series of different data collection

⁴⁵ For more information about selection bias, see: Results Measurement for Advisory Services (2007): Innovations in Impact Evaluation in IFC, [http://www.ifc.org/ifcext/rmas.nsf/AttachmentsByTitle/Innovationsmonitor/\\$FILE/Innovations2.pdf](http://www.ifc.org/ifcext/rmas.nsf/AttachmentsByTitle/Innovationsmonitor/$FILE/Innovations2.pdf)

methods. This might include using a regular survey as the core tool and combining it with in-depth discussions with a sample of those surveyed and interviewing key informants on particular key issues.

For example, in the Latvian business reform program FIAS employed a range of data collection techniques to assess and measure changes in the business environment. The main source of enterprise evaluation data was the ARCS. This was then combined with FG discussions, individual interviews and feedback sessions with both enterprises and government officials.

Case snapshot 4.6: Extending the enterprise survey in Lima

In Peru, IFC has helped the Municipality of Lima reform its business license procedures in order to cut the time, cost and number of requirements. The IFC Office for Advisory Services together with a local partner and MIT Poverty Action Lab designed and implemented a survey for the Lima Business Licensing Simplification project.

The quantitative analysis of core data about the number of days to obtain a license was complemented with qualitative interviews with entrepreneurs and key official of the municipality (both before and after the reform). Using these different sources of information allowed the evaluators to contrast the results obtained from different stakeholders and provide the full picture of the process. The evaluation demonstrated that the total cost of registration fell by more than 60%. The overall impact was an increase in registrations by 260% from the prior year.

> For further details, see the full case study on Lima in Annex 1

How should assessment criteria be applied to data?

Assessing project outputs and outcomes from the data that has been brought together during the evaluation process involves **analysis and judgment about benefits and success**. Such analysis typically involves a wide range of activities, including appraisal, assessment, examination, judgement, rating, reviewing, and testing. There are a number of techniques which can be used to facilitate this process. Two forms of assessment have been outlined as examples – performance scoring, and assessing cost effectiveness through quantitative analysis.

Performance scoring

Some organizations use scoring systems as an integral part of the review process to rate aspects of performance; for example, the likelihood that the outputs and outcomes of the project will succeed (or have succeeded, depending on when the scoring is done).

Annual scoring can provide important data for accountability, learning and decision-making. With care it may be possible for scores to be aggregated across a program or sector to provide an overall picture of success and value for money. ***The quality of scoring is clearly a key issue; since bad data will generate bad conclusions.*** The system has to be consistently and robustly applied involving relevant stakeholders and partners.

A typical scoring system uses a scale of 1-5 that can be applied for each output, for all outputs collectively, and at the outcome level. This is illustrated in Table 4.8.

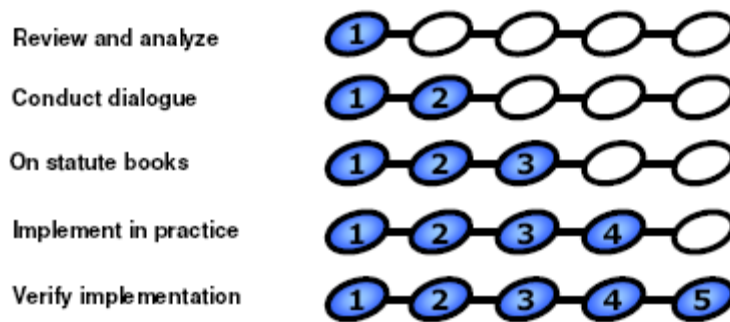
Table 4.8: Sample performance scorecard

No.	Descriptions	Achievement
1	Likely to be completely achieved	The outputs / outcome are well on the way to completion (or completed).
2	Likely to be largely achieved	There is good progress towards outcome completion and most outputs have been achieved, particularly the most important.
3	Likely to be partly achieved	Only partial achievement of the outcome is likely and/or achievement of some outputs.
4	Only likely to be achieved to a very limited extent	Very limited achievement of outcome and some outputs is likely.
5	Unlikely to be achieved	No progress on outputs or outcomes
X	Too early to judge	It is impossible to say whether there has been any progress towards the final achievement of outputs or outcome. This score should not be used unless they meet at least one of the following criteria: a) Postponement of project b) External constraints and or c) Recruitment delay

Such a scoring system could be used as part of a FG discussion with enterprises or government officials to help gauge their opinions about whether proposed changes in the regulations would be achieved.

Scoring systems are particularly useful for ‘process-oriented’ BEE interventions, such as regulatory governance or PPD initiatives. For example, PPD forums have been asked to assign a score from one to five to monitor government progress on reform proposals. This can be presented visually, as illustrated in figure 4.3.

Fig 4.3: Scorecard for government accountability



Case snapshot 4.7: The Latvian Business Environment Reform Program

Between 1999 and 2004 FIAS supported the Government of Latvia in the reform of inspections regulations and procedures. The initial reforms were enacted between 1999 and 2001 and it was expected that meaningful impact could only be discerned at the earliest in 2002 and more realistically in 2003 and beyond.

A critical driver underpinning M&E work in this program was the establishment of an overseeing Steering Group, appointed by the Prime Minister, with both public and private sector representatives. An 'Action Plan to Improve the Business Environment in Latvia' was developed as a legal instrument and updated regularly. The Steering Group have met on a regular basis since 1999 to review the Action Plan..

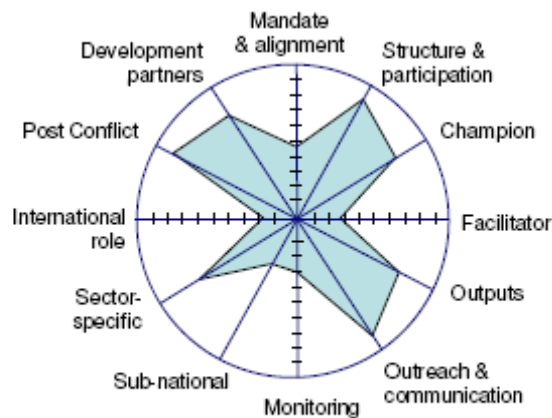
- Review and analyze proposals -achieved late 1998
- Conduct dialogue on draft reforms - achieved 1999
- Get reforms on the statute book – achieved early 2000
- Implement reforms in practice – achieved 2000/early 2001
- Verify implementation – 2001-2007- checked through ARCs

Administrative and Regulatory Cost Survey surveys (ARCs) conducted in 2001, 2003 2005 and 2007 provide ongoing evidence fro tracking and assessing performance.

> For further details, see the full case study on Latvia in Annex 1

In PPD reform, another useful tool – the evaluation wheel - has been developed to rate, analyse and present performance on 12 aspects of PDD activities (see figure 4.4). By plotting scores for each of these aspects along the spoke of the wheel, the 'shape' of performance for each dimension of PPD work can be observed and discussed. Each aspect on the wheel has associated indicators for measurement and a scoring system (from 0 = not satisfied to 5 = very satisfied) enabling the cross checking of data on similar aspects of the wheel.

Fig 4.4: Evaluation wheel for presenting performance of process indicators⁴⁶



The process indicators include scoring the existence of a mission statement and the ability to explain its content; the degree of participatory decision making; quality of management arrangements; quality and frequency of communication contribution made to conflict resolution; degree of autonomy from development partners. Therefore process indicators perform a function in relation to *how* the BEE is being reformed.

Assessing cost effectiveness through quantitative analysis.

Increasingly development partners are being asked to consider the cost effectiveness or efficiency of their interventions. Efficiency is an economic performance term comparing project outputs against the inputs. It illustrates the relation between means and ends and considers what extent the costs of a development intervention be justified by its results, taking into account alternatives; whether the intervention represents the quickest and/or cheapest way to transform investment into development gains, whilst minimising unnecessary transaction costs

Cost Benefit Analysis (CBA)

Cost benefit analysis (CBA) is a major evaluation instrument for projects with measurable benefits. For example, in business registration simplification, a CBA could consider whether the costs involved in providing technical assistance and support represent good value compared to the benefits gained through quicker and cheaper registration procedures.

⁴⁶ Source The PPD Handbook: A Toolkit for Business Environment Reformers 2006, p200

This raises the question of what standards to adopt as a reference point. The standard will sometimes be predetermined and will in other cases depend either on the terms of reference given to the evaluation team or the evaluator's own professional judgement.

In its simple form, CBA is carried out using only financial costs and financial benefits. For example, a simple cost benefit ratio for a road scheme would measure the cost of building the road, and compare this to the economic benefit of improving transport links. It would not measure either the cost of environmental damage or lower congestion or encouragement of new business activity attracted by improved transport links.

The CBA analysis depends on the timeframe of the costs and benefits being examined.

- Costs are either one-off, or may be ongoing.
- Benefits are most often received over time.

It is important to build this effect of time into the analysis by calculating the net present value including a discounted rate over time to reflect the opportunity cost of using resources.

CBA of a project or program can become an extremely complex exercise if all of the variables are considered, especially where the non-financial variables are many and difficult to quantify. A more sophisticated approach to building a cost benefit model is to try to put a financial value on intangible costs and benefits. This can be highly subjective. For example, an attempt to quantify the value to business of the reduced 'hassle' factor caused by bureaucratic delays.

Using CBA is not a new technique or tool for calculating efficiency but it is relatively new in terms of its application to BEE reforms.

Evaluation officers at IFC are working closely with advisory services project teams and business line leaders on a Cost-Benefit tool⁴⁷ to facilitate project reviews prior to approval; benchmarking/cross-project comparisons; project monitoring during project implementation.

However, the framework that has been developed so far by IFC is only for Business Registration projects similar frameworks should be developed for more BEE products.

⁴⁷ See the IFC Results Measurement website at <http://www.ifc.org/ifcext/rmas.nsf/Content/Home>

IFC is at the forefront of this work looking at the concept of 'private sector savings (PSS) or aggregate cost savings' as a universal impact indicator for business benefits arising from BEE reform interventions.

The concept of PSS is based upon calculating, estimating and extrapolating changes for a number of factors related to the reform interventions such as:

- Reduction in the direct costs to business - both formal and informal of starting (fees and time to get registration permits licenses etc) and operating a business (licenses, inspections, bribes etc) prior to and following reform; and
- Staff time saved due to the cut back in numbers of procedures and time taken with procedures

A number of larger BEE programs such as the regulatory reform in Latvia have used a CBA technique to examine the overall cost benefit impact of the reform program.

Case Snapshot 4.8: Using ex-post CBA in Latvia

Between 1998 and 2005, FIAS carried out six projects at a total cost less than \$500,000. The Government of Latvia also provided major in-kind contributions of staff time to oversee the reforms and covered most of the costs of the second and third business surveys. This amounted to roughly another \$500,000 putting the total cost of the work at roughly \$1 million.

Many of the FIAS recommendations were also supported by a World Bank loan for public administration reform (including tax and customs reform), which amounted to about \$45 million.

Taking the estimated benefit figure of US\$170 million between the period 2001 and 2005 gives a cost – benefit ratio of \$46 to \$170 (all discounted to 1998) meaning that each \$1 invested in the project resulted in at least \$3.7 savings for businesses in Latvia over a four-year period.

This shows that the project generated significantly more in monetary benefits than it cost to implement

Source: Source: FIAS Liepina et al 2006.

> For further details, see the full case study on Latvia in Annex 1

A different form of cost benefit quantification exercise can be undertaken using the results from an enterprise survey to estimate the saved costs to the average business, and from this extrapolating the total savings to the economy as a whole. In effect the economic impact.

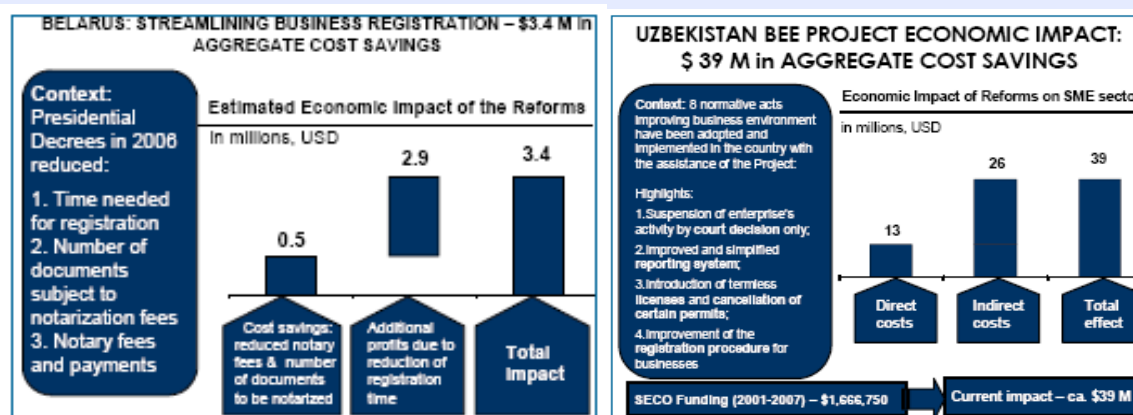
The methodology for doing this is described in detail in Annex 4.5 This methodology has been widely applied to BEE reform interventions in Eastern Europe and Central Asia as illustrated through the case snap shot below.

Case Snapshot 4.9: Using quantification techniques in Eastern Europe and Central Asia

The PEP IFC Advisory Services team has used the economic impact quantification technique to estimate an aggregate cost savings of US\$84 million for businesses in its focus countries. The calculations are made by comparing specific aspects of the business environment before and after IFC-supported reforms in order to quantify the benefits accruing to the target population – i.e. the aggregate savings to businesses.

In 2005 and 2006, the **Belarus** BEE project focused its efforts on simplifying costly and burdensome business registration procedures. As a result of the project's policy work, the government enacted a series of key changes that are estimated to result in direct cost savings to businesses of roughly \$500,000 (using the methodology described in Annex X). In addition, the shorter registration period is expected to generate about \$2.9m in profits to Belarusian SMEs. Thus the total expected economic impact is \$3.4m.

The SME survey conducted in early 2006 included specific questions on business experiences with the registration procedures. This allowed the project to capture a true pre-reform situation or 'baseline'. The project plans to conduct the next comprehensive SME survey in 2008 to capture the actual post-reform state of affairs. In the interim, in order to engage the government in a policy dialogue and estimate the impact of the reform, the project used expert assessments and official data to arrive at a conservative estimate of the aggregate cost savings to businesses. These will be verified once the data from the 2008 SME survey are available.



In **Uzbekistan**, since 2001 IFC has been working to improve the business environment by focusing on streamlining inspections, tax reporting, permit and licence issuance, and company registration procedures. Regular surveys allow tracking the impact of reforms over a longer period of time with clear pre- and post-reform benchmarks.

Inspections were the primary focus of the project's regulatory simplification work between 2002 and 2004. The SME enterprise survey conducted in 2001 established a baseline against which the project could track changes in actual business experiences with government inspections. Pervasive inspections were clearly one of the highest burdens for the private sector and represented a vehicle for extensive government intervention (and rent seeking) with no apparent benefit to the public. Substantial changes were progressively enacted with assistance of the project, resulting in streamlined inspections procedures and limited abuses as confirmed by the subsequent representative business surveys.

In Uzbekistan the effects of all the reforms that could be quantified were determined. The aggregate economic effect of eight Presidential decrees developed with in-depth assistance of IFC PEP experts during the life span of the project constitutes roughly US\$39 million for the SME sector. This consists of US\$13.4m in direct cost savings as a result of improved and streamlined inspection, permits, licencing, registration and reporting procedures. In addition,

these improved procedures are expected to generate approximately US\$25.9m in profits for SMEs in one year.

IFC is also applying this approach to other reforms enacted as a result of BEE project. In **Ukraine**, implementation of 2005 permits reform is still incomplete, and the inspections reform has just been adopted in 2007. However, it has been possible to calculate the burden of both procedures in terms of work time lost as a baseline for impact assessment. In addition, a first estimate of the savings resulting from the first phase of implementation of the permits reform (roughly US\$2m saved in 2006 as compared to 2004, and full implementation by the new regime of fire safety permits, allowing low-risk businesses to use self-certification, will result in savings of over \$31m for SMEs).

The most recent reform enacted in **Tajikistan**, the adoption of a far-reaching law on inspections in 2006, does not yet lend itself to full pre-post reform assessment, as its implementation is still in progress. IFC has calculated the overall burden of inspections, expressed as a percentage of the annual profits of businesses. This represents the baseline against which to measure the effect of the reform in a few year's time.

Source: Liepina, S, Dall'Olio, A & Sethi, S (2007): Smart Lessons: "Show me the money!" Quantifying the impact of regulatory simplification projects, IFC Smart Lessons in Advisory Services.

Undertaking CBA as part of BEE project evaluation can be useful but it is important to note that this technique has both advantages and limitations.

Advantages	Limitations.
<ul style="list-style-type: none"> ✓ A powerful, widely-used tool for estimating the efficiency of programs and projects. ✓ It can be used to help look at the ex-post impact of an intervention – did the investment generate the benefits (savings or returns) predicted or expected ✓ Can be useful tool for ex ante assessment when deciding whether to go forward with a project - does it look as if it will generate sufficient benefits to justify going ahead? ✓ Where costs or benefits are paid or received over time, it is possible to calculate the time it will take for the benefits to repay the costs. 	<ul style="list-style-type: none"> ✗ CBA can only be carried out reliably by using financial costs and financial benefits. If intangible items are included within the analysis an estimated value is required for these. This inevitably brings an element of subjectivity into the process. ✗ Fairly technical, requiring adequate financial and human resources. ✗ Requisite data for cost-benefit calculations may not be available, and projected results may be highly dependent on assumptions made. ✗ Results must be interpreted with care, particularly in projects where benefits are difficult to quantify.

What other resources are there on evaluation?

The above discussion presents some tools that are relevant for many BEE reform interventions. However there are a wide range of different data collection and assessment techniques and tools available for evaluation work. Table 4.8 below lists a number of key sources of information

Table 4.9: Key sources of information

Organisation	Web-link
World bank	http://www.worldbank.org/evaluation/
FAO	http://www.fao.org/pbe/pbee/en/about/index.html
IOD Parc: the Performance Assessment Resource Centre	http://www.parcinfo.org
IFAD	http://www.ifad.org/evaluation/guide/
EU Guidelines	http://europa.eu.int/comm/europeaid/evaluation
OECD and DAC	http://www.oecd.org/pages/ 0,2966,en_35038640_35039563_1_1_1_1_1,00.html
UNDP Evaluation Office	http://www.uneval.org/
International Development Evaluation Association	http://www.ideas-int.org/
IFC BEE toolkits	http://www.ifc.org/ifcext/sme.nsf/Content/BEE+Toolkits
IFC Results Management Advisory services	http://www.ifc.org/ifcext/rmas.nsf/Content/home

4.3 Assessing impact

What are the challenges?

Throughout the Handbook, the point is made that the pressure for 'proving impact' for aid interventions is increasing. It is now generally accepted that evaluation needs to evolve from its earlier focus on assessing outputs and outcomes to directly addressing impact. Development partners are increasingly seeking to improve their assessment approaches and techniques to help them make their impact findings robust, although there are methodological challenges to be overcome.

Some of the main methodological challenges concern issues surrounding the concept of causality and how to demonstrate the degree to which changes occurring in the BE can be **attributed** to a particular BEE reform intervention coupled with the ability to demonstrate that changes in the BE would not have happened if the BEE reform intervention had not taken place. This is known as the **counterfactual**.

This is the core of the **validation challenge** for measuring the impact. What are the strategies for overcoming this challenge? In general terms efforts can be made to tackle the validation challenge by ensuring that wherever possible three basic questions and principles of assessment are built into the evaluation design.

- i. **What was the situation before the intervention?** Provision of evidence for the project indicators are chosen **prior to, or at the beginning of the project**. Data collected at this time is normally referred to as '**baseline**' data and acts as the starting benchmark for the evaluation work. Baselines, as discussed previously in Section 3, are essential starting points and underpin all forms of effective impact assessment.
- ii. **What has happened after the intervention has occurred?** An ability to provide evidence relating to and on the output and outcome indicators chosen for key target beneficiaries of your project. This evidence when combined with the baseline will provide a basis by which directly comparisons can be made of the circumstances, experiences, attitudes and opinions of those to whom the BEE intervention is directed both **before and after** the intervention.

- iii. **What has happened because of the intervention?** An ability to assess whether impact has occurred **due to the intervention** requires some form of assessing results **‘with’** vis-à-vis **‘without’** the intervention. This is usually achieved by assigning some form of **control or comparator group** who have not had the opportunity to benefit from the intervention but whose situation/performance can be measured alongside the key beneficiaries of the project. These will be the comparator group and play a major part in helping to address the validation challenges of **attribution and the counterfactual**.

Different evaluation approaches with their associated methodologies make provision for attribution and the counterfactual to a greater or lesser extent. Three of the main approaches to evaluation given in Box 4.2 which also assesses the degree to which they help overcome these validation challenges.

Box 4.2: Evaluation Approaches

1. Non-experimental

Post-program judgment/expert opinion. (PPJ) Here the program participants are consulted after the intervention and asked to estimate the extent to which performance was enhanced as a direct result of the program

Before & After assessment (BAA). As the name suggests, this is a way to measure change by consulting with the program participants and measuring program indicators before (baseline data/information) and after receiving the intervention.

2. Quasi-experimental

These approaches compare intervention participants and some form of non-intervention control or comparator group both before and after the intervention. Different rationales are used to assign control groups but this is undertaken in a non randomised way.

3. Experimental

This approach looks at two groups before and after the intervention. There should be random assignment of the population into the project or treatment group who receive the intervention services and a control group, who do not.

For all three approaches, consideration should be given to:

- The underpinning principles of the approach and how it is used in practice.
- Its application, if any, to evaluating the impact of BEE reforms and
- The strengths and weaknesses of the approach vis-à-vis the other impact evaluation designs.

Experimental and quasi-experimental evaluation methods have typically been implemented in the areas of health and education, consequently, the application to BEE interventions are still in their ‘infancy’. The very nature of BEE projects and way in which they are undertaken, make it very challenging and at times impossible to apply these evaluation approaches in practice. IFC⁴⁸ is among the first to test out these methodologies in BEE reform. IFC’s Results Measurement Group is building a portfolio of evaluations that go beyond simple post and before-after comparisons. Evaluation designs are reviewed by experts and are implemented jointly with local partners, NGOs, universities, consulting firms and others within the World Bank Group. Many of the examples given below come from that work.

1. Non-experimental

These evaluation approaches are relatively easy to design methodologically, and are less expensive and complex to implement than experimental and quasi experimental designs. They are widely used in project and program evaluations, especially for smaller scale interventions. However, there are very few checks, if any, to address causality issues or to counter any potential bias in results arising from any sampling processes used.

Post-Program Judgment (PPJ)

Post-Program judgment (PPJ) is based on assessing the ‘after’ situation and is the simplest form of evaluation technically, the cheapest cost wise and hence is widely used.

PPJ is undertaken by examining the conditions and experiences of the key project stakeholders after the intervention activity has taken place. In this design, no baseline assessments are taken for the selected target individuals or groups. Impact evaluation is undertaken purely on the basis of measurements and assessments made *after* the intervention or activity has taken place. In this way the impact is measured on the basis of the stakeholders’ own understanding and reporting of the changes they have experienced both since and as a result of the intervention activity. There is no *a-priori* measure to act as a benchmark against which to compare the changes and experiences reported by the target group.

A key element for ensuring that the approach is as robust as possible is the use of rigorous sampling techniques in selecting relevant and representative subjects for the

⁴⁸ Monitor note ‘Innovations in Impact Evaluation in IFC’ Results Measurement for Advisory Services IFC <http://www.ifc.org/ifcext/rmas.nsf/Content/home>

evaluation exercise. Where possible the target groups should be selected randomly. For example, if a business simplification intervention is trying to improve the operating conditions for construction businesses in city 'A' then a sample of existing construction businesses who have been operating in city 'A' would be selected for the impact evaluation rather than printing businesses or construction business just starting in city 'B'.

Using PPJ for evaluating the impact of BEE reforms

PPJ approaches to impact assessment have been widely used for BEE interventions, particularly where there are limited resources and when the nature and context of the intervention means that there is no opportunity to undertake any form of 'causality checking' through examining a 'before and after' group of target stakeholders.

For example, in the case of business start up simplification it is not possible to create and before and after evaluation group for some key stakeholders. It is possible to take a sample of the government officials involved in the regulatory system and track their experiences and attitudes throughout and after the reform intervention. However businesses only go through the start up registration process once. Therefore it is not possible to ask this group about experiences 'before and after' the reform.

In these circumstances a simple post-reform evaluation is undertaken where the businesses that are going through the registration process post the reform program are asked what their experiences of the registration procedures and systems have been.

Whilst post interventions design and approaches are used for impact evaluation, in practice many make attempts to supplement the 'post experience' by reconstructing some form of 'before' comparative data for the evaluation exercise. While these measures cannot replicate a true 'before' situation insight into the effects of the intervention on the target groups can be achieved. Creating what are often termed **proxy 'baselines'** or **'before groups'** can be undertaken in several ways: using secondary data, from project records, recall and asking key informants. There are always strengths and limitations associated with any technique. Each technique is assessed below:

- **Secondary data.** Published information or other research papers give insight to conditions, circumstances, experiences and issues at the beginning of the project intervention.

Case Snapshot 4.10: IFC PEP MENA Alexandria business start up simplification project

This project used a several very detailed regulatory research reports produced by CIDA /SMEPol unit in the Ministry of Finance to help inform their baseline work and help estimate start up regulatory conditions practices and business experiences for a post project evaluation. In addition they have used their diagnostic survey work to help create a pre intervention business start up experience 'profile'.

They undertook a survey of businesses to complement their regulatory mapping work. This survey of 300 businesses who had registered in the three year period before the intervention began cannot act as a control group. However, their experiences do provide some form of pre start-up 'picture' including quantitative data for key indicators such as cost and time taken to register a business. This information will be used as a proxy baseline for the experiences of the business registering after the reform work has been completed.

> Further details are in the Egypt case study in Annex 1

Likewise information generated by other BEE reform interventions, especially research interventions can also be used to provide proxy before data. Here we illustrate three such BEE research sources: *'The Barometer 'Program de Mise á Niveau' in Tunisia'* has been providing economic climate measurement and competitiveness data based on enterprise experiences on an annual basis since the late 1990s; the Philippines City Simplification Survey and in Mongolia the annual export Climate survey provides regular information for this area of the BEE.

Case Snapshot 4.11: The Barometer 'Program de Mise á Niveau' in Tunisia

In Tunisia, since 1998 GTZ has been working with the Tunisian Ministry for Economy and Energy under the framework of the national economy promotion "Program Mise a Niveau" (PMN). The program focuses on the improvement of the environment for business including administration, transport and infrastructure, and also supports industry and related service providers.

GTZ, in partnership with IFO-institute Munich developed a qualitative measurement instrument for economic climate measurement instruments. The 'barometer' works through a regular survey of representative samples of enterprises participating in PMN in order to reach conclusions about whether the participation contributed to increasing their competitiveness and thus measuring the efficiency of the program.

The sampling methodology is based on a quota which is determined according to defined characteristics. A tracking approach was adopted where each sample is newly selected from the total unit for each set of enquiries which occur on a 4-monthly basis. The assessment is done as a self-administered enquiry (without an interviewer) by delivery of a questionnaire by fax.

To measure competitiveness directly is virtually impossible as it is determined by a multitude of single factors which are specific to the respective enterprise and market. The questionnaire consists of 10 questions, 7 of which are standard and 3 of which are variable and can be

tailored to suit a specific category or situation. The structure follows a qualitative approach based on consumer confidence. Rather than asking how much money has been invested in certain forms of assets, it asks whether the enterprise is running better than 2 years ago.

Source: Detlev Jahn, Project Coordinator, Private Sector Promotion, GTZ Tunisia

Case Snapshot 4.12: Philippines City Simplification Survey

The Cities Competitiveness Ratings project (PCCRP) is the flagship M&E tool for the enabling environment component of the GTZ Small and Medium Enterprise Development for Sustainable Employment Program (SMEDSEP) in the Philippines. The survey is the result of collaboration between SMEDSEP and the Asian Institute of Management (AIM) Policy Centre.

The objective of the survey is to benchmark the competitiveness of cities to allow comparison among them and over time. The survey measures seven 'drivers of competitiveness' to identify strengths and weaknesses: dynamism of local economy, human resources and training, responsiveness of local government, quality of life, infrastructure, linkages and accessibility, cost of doing business. Each driver encompasses 70 qualitative and quantitative indicators.

In each participating city, a partnership was formed with a local academy to implement the survey and the overall effort was coordinated by a well established academic authority in the country. A major advantage of this data source is that it is an objective source of information and the results cannot be influenced by the program.

> More detail is provided in Annex 1: Case Study on Philippines

Source: Smart Lessons, GTZ SMEDSEP Monitoring Manual

Case Snapshot 4.13: The IFO Export Climate Survey in Mongolia

The GTZ funded Export Climate Survey, Mongolia, has been developed by experts from the IFO institute for Economic Research. Conducted on a yearly basis, it covers companies operating in mining, manufacturing, tourism, transport and trade sectors.

The monitoring of export-oriented companies aims at identifying the most important obstacles to exporting as seen by entrepreneurs. Rather than providing a 'one shot in time' static picture, it is designed to show the process of change over the years by replication with the same sample of entrepreneurs every 3 months.

The methodology is based on consecutive (periodical) qualitative surveys. The questions are not designed to collect precise figures, but rather provide information on opinions and directions of change on the importance of obstacles to exporting and export conditions. As no precise quantitative figures are generated, the preferred statistical method is not the selection of a random sample for each survey, but to build up a panel of respondents that remains relatively consistent over the course of the survey period.

Source: GTZ (2005, 2006): Series on Industrial and Trade Policy, Export Climate Survey Mongolia

- **Project records** from the intervention itself; other projects in the development partner/agency; the work of other development partners. All can be used to help build a picture of the pre intervention context.
- **Recall** can be used to estimate conditions prior to the project. While recall is generally agreed not to be a reliable way to obtain precise numerical data such as financial data, it may be a valid way to obtain information on major changes or events. Care should be taken to avoid bias when using such information.
- **Key informants.** These are people who have some particular or in depth knowledge or experience. For BEE reform, business leaders, business associations, local government agencies, researchers support business support professionals etc. may be regarded as key informants be able to provide useful reference data on baseline conditions. Again caution should be used as some such sources may be biased.

The above sources of information are not ideal but using several of them to triangulate findings can help to build a relatively credible picture.

Using PPJ as a means of assessing impact for BEE interventions has a number of strengths and limitations (see table 4.9), however overall it is the least robust method of evaluation with no 'causality' checks for validating results.

Table 4.9: Strengths and weakness of post-project judgment evaluation

Post Project Judgment Evaluation for BEE Reform	
Strengths	Limitations
<ul style="list-style-type: none"> ✓ It is low cost compared to other designs ✓ Often the only option available when there are data and budget constraints. ✓ The design captures data on change at only one point and so is easier to conduct than having to identify and select control groups. ✓ Several BEE programs have been evaluated utilising this approach and so there is practical experience to draw upon. 	<ul style="list-style-type: none"> ✗ This approach relies on program participants or independent experts to make judgments concerning impacts with no control for the counterfactual. ✗ Care needs to be taken to make sure that people consider the counterfactual in their assessment of impacts. ✗ The design does not attempt to <u>understand</u> any changes that have occurred and assumes that they have occurred as a result of the BEE reform ✗ Does not capture process issues from the reform implementation

‘Before and After’ Assessment (BAA)

BAA in practice

As the name suggests, a ‘Before and After Assessment’ examines the experiences and circumstances of a given target group of target stakeholders both before and after they have experienced the intervention using a selection of indicators.

The aim is to establish if any changes in the indicator criteria have taken place for the identified target group. These changes in the indicator criteria are then analyzed in order to determine the impact of the intervention.

A key element for ensuring that this approach is as robust as possible is the **use of rigorous sampling techniques**. Ideally the target groups for the evaluation should be selected randomly and within the parameters of the specific stakeholder population. The target groups selected for BAA must be:

- **Relevant to the intervention being examined:** they must come from those individuals and groups who are key stakeholders for the intervention activity being evaluated. In BEE reforms typical sample groups will be businesses and government officials
- **Representative of the key stakeholder population:** they should be the type of individuals or groups that are directly involved in and/or likely to be affected by the intervention activity being evaluated. In BEE start up reforms typical sample groups will be: new businesses, business operating informally that are now formalizing and government officials who are involved with this area of activity be this at policy or an operational level. If interventions apply to a specific location or a specific sector then only participants’ from these areas and or sectors will be considered for selection.
- **Representative of any diversity within the key stakeholder population:** if the target group is very diverse in terms of its characteristics – age / size / gender / location etc. – it may be necessary to ensure that a proportion of groups or individuals from each of these sub groups are represented within the sample selected. This is known as **stratified sampling**. If the intervention is being undertaken throughout an area with distinct sub districts where conditions relating to the area vary, then it would be important to ensure that the sample group selected included representatives from these different groupings.

Taking these sampling factors into account and establishing a relevant and representative set of individuals or groups will also help to determine the total numbers to be included in the evaluation group.

Using BAA for evaluating the impact of BEE reforms

The evaluation design has been widely adopted in many BEE reforms, particularly where it is possible to identify and measure the benefits that come about as a result of the reform interventions, for example, reduced time and cost spent starting a business or complying with government inspections business

Case snapshot 4.14: Latvia inspections reform

Between 1999 and 2004 FIAS supported the Government of Latvia in the reform of inspections regulations and procedure which resulted in the following reforms

- Reduced inspection burden on businesses, including incidence of inspections from labor, sanitary, construction and municipal police and a shorter average duration of inspections at one enterprise.
- Development and implementation of a compliance-oriented approach in the inspectorates
- Improved positioning of legal, procedural and technical information to businesses by the inspectorates via brochures, websites and training seminars
- Greater quality and professionalism of inspectors

The reforms were enacted between 1999 and 2001 and it was expected that meaningful impact could only be discerned at the earliest in 2002 and more realistically in 2003 and beyond. ARCS surveys conducted in 2001, 2002 and 2003 provided time-series data which captured data on inspections. These business surveys confirmed that the average frequency and duration of inspections for the five most common inspections (fire safety, labor, construction, environmental and sanitary) fell between 2001 and 2003, resulting in a saving of around 39 hours per year for an average firm.

Source: FIAS (2006): pp10-11

> For further details see the full case study on Latvia in Annex 1

Case snapshot 4.15: Lima City Simplification

In Peru, IFC has helped the Municipality of Lima reform its business license procedures in order to cut the time, cost and number of requirements. The survey forms part of an evaluation using a Before and After methodology interviewing a sample of owners of newly licensed businesses before the reform as well as owners of newly licensed businesses after.

The team conducted three rounds of interviews of 50 firms each, two before the reform (August 2005, October 2005) and one after the reform (September 2006). The two pre-reform rounds were designed to check that there were not significantly different results in terms of number of visits, length of time, number of requirements or the cost, before the reforms. This confirmed that in the absence of reform there was very little change before the reforms were introduced.

Source: Smart Lessons in Advisory Services: How the project evaluation results don't just go to a shelf. Business licensing simplification in Lima, Peru

> For further details see the full case study on Peru in Annex 1

Using the BAA as an evaluation methodology for impact assessment of BEE interventions provides some attempt to objectively assess the changes experienced by the target groups of the intervention. At the national level using control groups is problematic, since all businesses will be subject to the new reforms, there is hence no identifiable control group against which to measure. Thus in this case, this is not a particularly robust methodology in terms of validating impact results.

However, at the sub-national level, this technique is possible if a control group can be identified in a similar location (city, municipality, province) where the reforms have not been applied. It is important that control location have similar profiles to the target locations to ensure that other factors remain equal (as far as this is possible).

Alternatively, it could be done if the simplified procedure is being rolled out as a pilot so that control and treatment groups can be identified. It should be noted that the ethical and political considerations of undertaking this type of study make it challenging.

Table 4.10: Strengths and weakness for BAA

Before and After Assessment for BEE Reform	
Strengths	Limitations
<ul style="list-style-type: none"> ✓ This design attempts to capture and understand any changes that have occurred rather than assuming that they have occurred. ✓ Individuals are asked to estimate the extent to which performance was enhanced as a direct result of the program – in effect, to compare current performance to what would have happened in the absence of the program ✓ Working with the same group is cheaper than identifying and selecting control groups which is often simply not possible. . ✓ Several BEE programs have been evaluated utilising this approach and so there is practical experience to draw upon. 	<ul style="list-style-type: none"> ✗ The design cannot isolate the impact of the program from extraneous factors such as selection bias, maturational trends, secular drift and interfering events. ✗ This approach relies on program participants or independent experts to make judgments concerning impacts. ✗ This approach requires people to be able to determine the net effect of the intervention based solely on their own knowledge and experience

2. Quasi-Experimental Designs (QEDs)

In QED approaches, explicit attempts are made to address the validation challenges of attribution and the counterfactual when evaluating the impact of an intervention.

This is achieved by setting out to examine changes experienced by the project target group (sometimes called the 'treatment group') i.e. those 'experiencing' the intervention, and comparing them to a set of people 'not experiencing' the intervention. This is usually tackled by assigning some form of control or comparator group i.e. a group who have not had the opportunity to benefit from the intervention but whose characteristics are similar to those that have, and whose situation/performance can be measured alongside the key beneficiaries of the project.

A control or comparator group is created or selected that is composed in a non-random way, but provides the counterfactual to a 'treatment group'. To the extent that the two groups are similar, observed differences can be attributed to the BEE intervention being evaluated with a higher degree of confidence than in the simpler PPJ and BAA approaches.

Several methodologies are used for creating control or comparator groups. One of the most widely used is that of **matched comparisons**. Matching involves identifying non-project/program participants comparable in the essential characteristics to participants. Both groups should be matched on the basis of either a few observed characteristics or a larger number of characteristics that are known or believed to influence program outcomes.

In practice, it is rarely possible to construct a 100% perfectly matched control group, or even to measure all possible relevant characteristics. Nevertheless, matching can be achieved for key characteristics and this is widely regarded as a rigorous methodology when evidence is available to show that treatment and control groups are similar enough to produce a close approximation to the perfect match.

Using quasi-experimental design for BEE reforms

With BEE reforms it can be difficult to find matched groups because of the need to find groups not exposed to the reform intervention that are similar in key characteristics to those that are involved in the reform. BEE interventions by their very purpose, improving the BE, are universal in nature and hence apply to all groups or are sector specific. In the former case it would be unethical and a violation of a governments remit to be

discriminatory if it were to limited access to reforms to selected businesses. In the latter case selecting businesses outside of the sector would not result in reasonable matches.

While recognizing these challenges, the matched comparison QED approach is increasing being tested out for BEE interventions as the Balkans case snapshots show.

Case snapshot 4.16: Using matched comparisons for ADR reform in the Balkans

In 2006, an independent evaluator looked at the experiences of two groups in two pilot cities (Banja Luka in Bosnia and Herzegovina and Belgrade in Serbia) through two surveys.

In selecting the two groups of business the main criteria were that:

- Members of both groups need to have accepted the offer of mediation initially. This was to ensure that their case has passed the criteria for referral to mediation and that the individuals were open to an alternative ways of resolving their dispute.
- There should be a good mix of plaintiffs and defendants in each group.
- The nature of the dispute should be sufficiently comparable ie commercial, civil, labor

The first condition was important to enable meaningful comparison of these two groups

- The treatment group (known as the Quasi Experimental Group 'QEG') were those who had utilized mediation after 3 to 6. This group comprised 155 businesses in Belgrade and 142 in Banja Luka; and
- The control group (CG) were those who did not utilize mediation because, on second thoughts, they rejected the proposal or because one or both of the parties did not come to a scheduled mediation within 6- 12 months (Control Group 'CG'). This group comprised 71 businesses in Belgrade and 70 in Banja Luka

The report first looked at the 'matching' of the two groups in each country in terms of:

- Type of dispute
- Company profile of staff, sector, turnover, legal department
- Role in dispute
- Nature of dispute
- Size of dispute
- Attempts to resolve prior to mediation
- Results of mediation

The evaluator then looked at a series of questions that related to key indicators:

- Duration of case
- Cost of case
- Value of dispute compared to value of settlement
- Enforcement or fulfillment of dispute

As a result, there were some very useful insights achieved. The evaluator went on to make a number of recommendations on how to improve the comparability or matching of the groups that has informed the development of this technique. The differences in the legal framework in the two countries is less significant in terms of creating matching comparisons than the matching of the two groups in each country to enable meaningful and rigorous analysis.

> More detail is provided in the Case Study on Balkans presented in Annex 1.

Using QEDs as an evaluation approach for impact assessment of BEE interventions provides a robust attempt to objectively assess the changes experienced by those for whom the intervention is intended to benefit. Introducing control groups both before and after the intervention into the evaluation provides a comparator against which to examine issues of causality. The strength of QED approaches depend on the rigor with which the control groups are matched to the target group being evaluated and in practice getting a good match is difficult.

For example, if the regulatory intervention being evaluated is business registration, since this is a one-off procedure, the control group (those registering under the old regime) and the target group (those registering through the revised regulatory process) will necessarily be different groups of businesses. Care must therefore be taken to create a sample of both groups that have similar characteristics (for example, firm size and sector). If the regulatory intervention is affecting something that must be renewed or is undertaken annually or more frequently (for example, an operating license, a business inspection, a health and safety assessment), then it may be possible to use the same sample of firms. However, it should be noted that while this may be successful if the sample is of large, or possibly medium sized firms, it is unlikely to be effective for a sample of small or micro firms where rates of market exit is high.

However the application of such approaches to some BEE intervention is attempting to address the challenges of validating impact assessment and bringing more robust evaluation methodologies to the sector.

Using QED approaches for impact assessment of BEE reform has a number of strengths and limitations (Table 4.11).

Table 4.11: Strengths and weaknesses of quasi-experimental designs

Quasi Experimental designs for BEE Reform	
Strengths	Limitations
<ul style="list-style-type: none"> ✓ These designs face less of the ethical or political problems of excluding groups from the reforms and their benefits. ✓ They can often draw on existing data sources and are thus often quicker and cheaper to implement. ✓ They are well used in practice albeit outside of the PSD /BEE reform field. ✓ Matching is a relatively easy process compared to randomized allocation ✓ There are a variety of methods to use in generating or selecting comparator groups depending on the nature of the activity being evaluated. 	<ul style="list-style-type: none"> ✗ The reliability of results is highly dependent on the 'matching methods' which can be difficult to conduct. ✗ Valid comparisons require that the two groups be similar with respect to key characteristics, exposure to external events and trends, and propensity for program participation. This can be difficult to establish. ✗ Because the two groups are essentially 'non-equivalent', the possibility exists that at least some of the differences in outcomes may be explained by unobserved variables that differ across the two groups. ✗ Requires considerable expertise in the design of the evaluation and in analysis and interpretation of the results. ✗ There is little practice with BEE as yet.

3. Experimental Designs⁴⁹

Bias can occur for a host of reasons and take many different forms. For example, sampling bias occurs in the selection of target groups when only those who have offices within a short distance of the one stop shop are included. As noted earlier in this section, practical attempts are made to mitigate this bias by the hiring of external experts who are not connected with the project and have the technical expertise to ensure that appropriate methodology design and sampling is conducted. However some would argue that the only robust way of tackling bias is by using *experimental designs* in evaluations. Randomization is a key feature of experimental approaches. This is considered the most rigorous of the evaluation methodologies, the 'gold standard' in

⁴⁹ Based on a note 'The Encouragement Design for Program Evaluation 17 September 2007 Alexis Diamond (IFC & Harvard University) Jens Hainmueller (Harvard University & IFC) [http://www.ifc.org/ifcext/rmas.nsf/AttachmentsByTitle/Encouragement/\\$FILE/The+Encouragement+Design+for+Program+Evaluation.pdf](http://www.ifc.org/ifcext/rmas.nsf/AttachmentsByTitle/Encouragement/$FILE/The+Encouragement+Design+for+Program+Evaluation.pdf)

evaluation. This is especially the case when we are trying to estimate the effect of an intervention on a complex concept of the BEE.

In a randomized experiment, the researcher cannot manipulate the group who are 'exposed' to the intervention (the "treatment group") and the not-exposed group (the "control group"). Randomization ensures that, on average, prior to the intervention, treatment and control groups are essentially identical and therefore would show very similar results in the absence of the treatment. Therefore, a difference in results for the two groups can be causally attributed to the program. This design copes with the challenge of attribution and the counterfactual.

Using experimental designs for BEE reform

In practice, randomised experimental designs have not been used in BEE type of work for a number of reasons:

- BEE type of interventions, take place in 'open systems' where multiple players are operating trying to bring about change and there are numerous stakeholders involved as implementer's and beneficiaries as well as the evaluators themselves. Therefore attempts to control the environment in which the intervention takes place is nigh on impossible.
- It is difficult to ensure that all those selected for treatment get 'the treatment' as such and all of those selected for control group do not. Many BEE interventions are not so simple, specific and direct that their effect can be measured in the same way as say a vaccination shot to a child can be measured in a health study.
- It is sometimes impractical and could be regarded as unethical to force or encourage some of the subjects to be in the control group. Excluding businesses from interventions that are purposely aimed at improving their operating environment would seem perverse and go against creating a more level playing field for all. For example in a business-simplification reform we would like to estimate the effect of registering a business on firm-level impacts, but we cannot force some firms to become registered and other firms to remain informal.

Therefore while experimental designs are seen as the 'gold standard' their practical application for impact assessment of BEE reforms is only at an exploratory stage largely because of the technical challenges that limit their use in this field of work as shown in table 4.12.

Table 4.12: Strengths and Weaknesses for experimental randomized designs

Experimental Randomized designs for BEE Reform	
Strengths	Limitations
<ul style="list-style-type: none"> ✓ Random assignment helps guarantee that the two sample groups are similar. ✓ Extraneous factors that influence outcomes are present in both groups. ✓ Because of this comparability, claims that differences between the two groups are the direct result of the program are more difficult to refute. Interpreting the results is simple. ✓ Experimental designs are used extensively to test the efficacy of new treatments in health, social welfare and education. 	<ul style="list-style-type: none"> ✗ Denial of assistance to some is seen as unethical. ✗ It can be politically difficult to provide an intervention to one group and not another. ✗ The scope of many BEE reforms are nationwide programs or policy changes which rule out the possibility of selecting a control group although encouragement design can potentially help address this. ✗ It may be difficult to avoid selection bias and ensure that assignment of treatment and control groups are truly random. ✗ It takes significant planning and management to ensure that the services provided to both entities are exactly the same. ✗ Experimental designs can be expensive and are time consuming. ✗ Requires high level evaluation skills. ✗ There is little practice with BEE reforms from which to draw upon.

What is the best approach?

The reality of current practice in assessing the impact of BEE reform interventions is that there is much wider practice of simple post program judgment and before and after approaches than quasi experimental approaches. Efforts are being made, with strong leadership from the IFC Results Measurement team, to improve awareness of and the technical capability for applying QED approaches to evaluation work. This Handbook

along with other resources reflects this movement to ‘upgrade’ the rigor of evaluation for BEE interventions.

Table 4.13: Summary of key characteristics for different evaluation approaches for impact

Evaluation activity	Post Program Judgement	Before and After	Quasi Experimental	Experimental
Post project assessment	✓	✓	✓	✓
Before project assessment	✗	✓	✓	✓
Use of target groups	✓	✓	✓	✓
Use of control groups	✗	✗	✓	✓
Use of randomly selected groups	✗	✗	✗	✓
Level of technical skills needed to design	Low	Medium	High	Very High
Cost of undertaking	Low	medium	High	Very high

Table 4.14: What types of impact assessments are appropriate for regulatory simplification?

	National level reform		Sub-national reform	
	Business registration	Business operations	Business registration	Business operations
Economic impact quantification using enterprise survey data	✓	✓	✓	✓
Before and After (sample of companies is not constant)	✓	✓	✓	✓
Before and After (sample of companies remains at least partly constant)		✓		✓
Quasi experimental with some form of control comparator	X	X	✓	✓

4.4 Forthcoming developments in M&E for BEE

Practice in M&E for BEE interventions is currently being developed rapidly and new techniques and tools being developed all the time. Measurement, quantification and evidence-based policy making are becoming increasingly dominant features in the approach of many countries.

The latest two BEE toolkits have significant sections on M&E issues, namely the PPD Handbook and the Strategic Communications for BEE Reforms Toolkits..

Evaluation groups in many donor and development organizations are also working on further developing good practice. For example, the Results Measurement Unit in IFC is developing a range of easy to use evaluation tools. Currently in development is a standardized methodology for CBA, and guidance for embedding M&E more actively within project appraisal forms in IFCs DOTs project management systems. For updates on this work, refer to the Results Measurement Unit website at: <http://www.ifc.org/ifcext/rmas.nsf/Content/home>

In addition a number of IFC offices are developing pilot studies using control groups and some quasi-experimental methodologies, many of which have profiled in this Handbook and its accompanying annexes.

At the time of writing, the DCED are finalizing their donor guidelines on BEE interventions entitled “Supporting Business Environment Reforms: Practical Guidance for Development Agencies”. The guidance advocates 4 phases to BE reform, namely: diagnostics, design, implementation, and monitoring and evaluation. It is emphasized that a sound M&E system is essential for the success of BE reform programs including well-defined indicators that measure outputs and outcomes and clearly connect outputs with outcomes and their impact on poverty. The guidance also notes that special attention should be given to assessing the impact of BE reform on enterprise development and pro-poor growth. It also highlights that measuring this impact should be a partnership between the development agency, the government and the private sector. Following the publication of this guidance, the DCED intend to commission further work on developing standards and guidance specifically on evaluation and impact assessment.

This forthcoming work from the DCED is likely to draw on attempts to standardize methodologies for national measurements so that comparisons can be made across

countries. To date, this work has been pioneered by the OECD's Regulatory Reform group and the Sigma program. A number of development partners are now looking to customize and apply similar techniques to developing countries.

In summary the issues of monitoring evaluation and assessing impact for BEE reforms is a hive of development and debate. This Handbook presents a resource that brings together examples from current practice in order to help raise awareness, engage interest and improve good practice across all BEE reform interventions

4.5 Key messages

- The imperative to improve development results has generated a demand for the effective evaluation of BEE reforms.
- Evaluation can take place alongside project design and appraisal - it is not exclusively an ex post activity
- Who undertakes evaluation is an important consideration and can affect levels of inclusion and diversity.
- A distinction can be drawn between review evaluations and assessing impact based on the timing, focus and then related to the type of 'results' achieved
- There are essentially three tasks: which questions; what data/information and what success criteria to employ
- The compilation of good quality baselines are critical for meaningful impact assessment and must be produced wherever possible.
- Experience and practice is growing and innovative approaches are being tried. The honest sharing of experience will improve the ability to undertake evaluations for BEE interventions
- The adoption of robust impact designs and methodologies is essential in order to address the validation challenges of attribution and the counterfactual. Truly experimental designs are difficult to achieve in BEE work.
- While investment and economic growth are the primary indicators of BEE reform success, social inclusion and poverty alleviation considerations will affect long term sustainability. Improving the integration of equity and sustainability issues is critical to the broader understanding of impact.