



THE PRESIDENCY
REPUBLIC OF SOUTH AFRICA

DEPARTMENT: PERFORMANCE MONITORING AND EVALUATION

DPME Evaluation Guideline No 2.2.13 Guideline on Impact Evaluation

Created 20 March 2014

Addressed to	Government departments who are undertaking evaluations (programme managers and M&E staff) as well as evaluators of government programmes and policies.
Purpose	The purpose of this Guideline is to provide technical guidance on undertaking and managing a Impact Evaluation
Policy reference	This guideline should be read in conjunction with the National Evaluation Policy Framework approved by Cabinet on 23 November 2011 (available on the DPME website).
Contact person for this guideline	Jabu Mathe, Evaluation and Research Unit (ERU) E-mail: jabu@po-dpme.gov.za Tel: 012 312 0158

1 Introduction

Impact evaluations seek to understand the changes brought about by an intervention, at outcome and impact level. This Guideline is designed to assist government departments to effectively plan and manage an impact evaluation. The Guideline provides a definition and description of impact evaluation followed by key questions impact evaluations can answer, guidance on key forms of impact evaluation and its common methods and approaches, key issues to be considered, and how to manage the evaluation. These are broad guidelines that can be applied in different contexts. It is focused on providing an overview for government staff managing evaluations and is not targeted as a manual for an evaluator on how to undertake an impact evaluation. Hence it does not go into detail into the different tools that are introduced. Note the word programme is used here but the evaluation could equally apply to a policy, plan or project – we use intervention to cover any of these.

2 Definition of impact evaluation

Figure 1 shows the results-based management pyramid, which has the logic chain of activities resulting in outputs, these resulting in outcomes being achieved and finally impacts. For example, achieving the outcomes of improved access to land and increased levels of participation in community decision-making might occur before, and contribute to, the intended final impact of improved health and well-being for women. The distinction between outcomes and impacts can be relative, and depends on the stated objectives of an intervention (betterevaluation.org).

The Asian Development Bank guidelines state that impact evaluation establishes whether the intervention had a welfare effect on individuals, households, and communities, and whether this effect can be attributed to the concerned intervention.¹

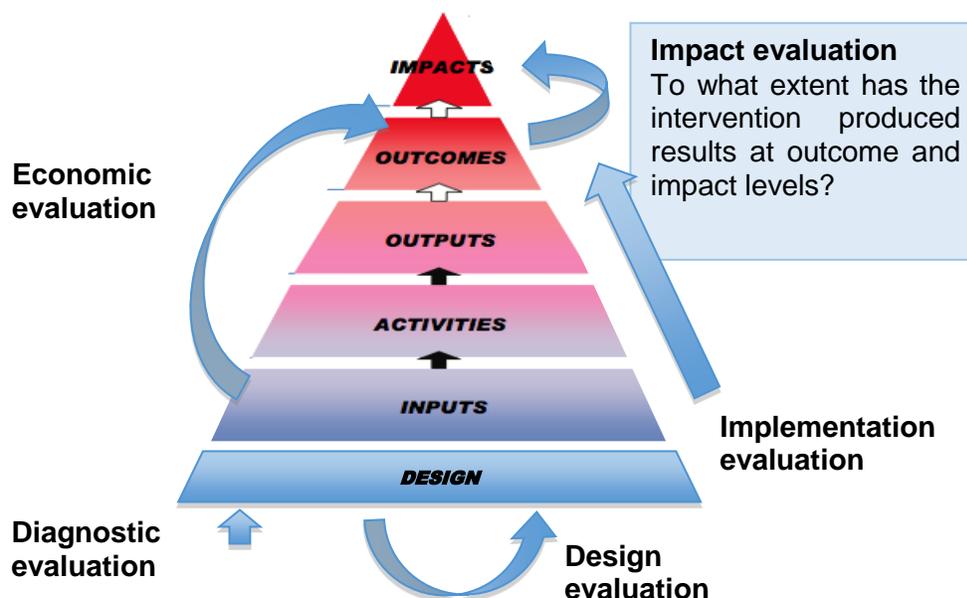
¹ <http://www.adb.org/data/economic-research-initiatives/implementing-impact-evaluation-at-adb>

This works for interventions with have quite close linkages with services to people, but for lower level interventions there impacts may be at system level.

An impact evaluation needs information about impacts – but also information about inputs, activities, outputs and outcomes. Also, for policy making purposes it requires information about costs. It is therefore complemented by other types of evaluation found in the DPME framework. In some cases, where the linkages between the output and outcome are direct and clearly established, impact evaluations may just examine the extent to which outputs have been achieved. So for example, *correct* use of malaria bednets in a programme directed at reducing the incidence of Malaria, or, the uptake of male circumcision in a programme directed at reducing the incidence of HIV/AIDs are all legitimate impact evaluations.

Ideally an impact evaluation should be designed prior to implementation of the intervention and conducted late enough for impacts, or longer-term outcomes, to be evident. The theory of change that underpins a programme or project can identify key points when it will be useful to collect data for an impact evaluation. Usually as well as trying to understand the impacts, one also wants to understand why and how they occurred, and how positive impacts could be strengthened, and negative impacts lessened. This latter element requires the methodologies of an implementation evaluation. Therefore in most impact evaluations, it is important to think about a process evaluation and process monitoring as an integral part of the impact evaluation. Thus for example 3ie's impact evaluation of cash transfer programmes, included process evaluations. These helped the government understand the need for better training and better targeting *before* the impact evaluation concluded. This attribute of impact evaluations can help in better programme management as well.

Figure 1: Relationship of evaluations to results-based management



In rare cases only the impact is wanted (such as the Impact Evaluation of Grade R carried out in 2013), which asked does Grade R result in improved learning outcomes (i.e. have the impact that was anticipated), but in most cases it is both, and an implementation evaluation will be

carried out at the same time. In many cases too questions of cost-effectiveness will be raised, and so a component will also be an economic evaluation.

3 Purpose of impact evaluations

The National Evaluation Policy Framework (NEPF, 2011) describes impact evaluation as an approach that measures changes in outcomes, and the well-being of target beneficiaries, that are attributable to a specific intervention. Impact evaluations assess the causal links between an intervention and identified changes, usually comparing with a counterfactual (what would have happened if the intervention had not happened).

In this guidance note, an impact evaluation includes any evaluation that systematically and empirically investigates the changes produced by an intervention, whether at outcome or impact level and whether the policy, programme or project under investigation was responsible for this impact. Note that in all cases it should be combined with an implementation evaluation to understand if the intervention is working as designed or if there are important links in the causal pathway that are not being realized (uptake, behaviour change, adoption, delivery of inputs etc.).

Impact evaluations often assess the effectiveness of a policy (does it work?), the efficiency of resources (value for money) and adaptability – indicating when and how to modify the policy/programme.

This serves four different purposes:

- **Informing policy decisions** – the impact evaluation of large programmes, or the inclusion of impact evaluation data and findings in synthesis evaluations, can provide useful and convincing evidence to support decisions on public policy, including deciding which programmes will be funded in the future. Cost effectiveness and cost-utility comparisons can help compare different policy options, and as impact evaluations are measuring effectiveness, are usually possible.
- **Improving intervention design and implementation** – impact evaluations that can show impact but also explain how programs and projects work, and what is needed to make them work well, can inform and improve the design of future similar interventions. In this case they will combine with an implementation evaluation (see Guideline 2.2.12 on Implementation Evaluation).
- **Accountability** – Impact evaluation of government policies and programmes shows whether public funds are making a difference, and the extent to which the public interest has been effectively served. Even where an impact evaluation finds that a programme or policy has not worked, the results can be used to improve the allocation of future resources, improving accountability.
- **Informing delivery**: Impact evaluations can be very useful in comparing different ways to deliver and implement a policy. In most development programmes, it is often delivery and implementation that requires information and informs the effectiveness of interventions rather than their efficacy.

It is particularly important to do impact evaluations for major schemes where government is investing large amounts of money, innovative schemes, pilot programmes which are due to be substantially scaled up, interventions for which there is scant solid evidence of impact in the given context, or a selection of other interventions across an agency's portfolio on an occasional basis, etc.

4 Typical questions

Examples of evaluation questions and subquestions for impact evaluation include (adapted from Rogers, 2012):

What was the overall impact of the intervention?

- Did the intervention (programme, project or policy) work? Did the intervention produce the intended impacts in the short, medium and long term?
- Was the impact attributable to the policy/programme under review?
- For whom, in what ways and in what circumstances did the intervention work?
- How *much* did the intended beneficiaries benefit and to what extent did the impacts meet their needs?
- What unintended impacts (positive and negative) did the intervention produce?
- Much broader - is this the best intervention to achieve the desired outcome?

What is the nature of the impacts and their distribution?

- Are impacts likely to be sustainable and durable?
- Did these impacts reach all intended beneficiaries? If not, why not?

What other factors have influenced the intervention to achieve impact?

- How did the intervention work in conjunction with other interventions, programmes or services to achieve outcomes?
- What helped or hindered the intervention to achieve these impacts?

How did the intervention work to achieve (or not to achieve) impact?²

- How did the intervention contribute to the intended impacts?
- What were the particular features of the intervention that made a difference?
- What variations were there in implementation?
- What has been the quality of implementation in different sites?
- To what extent are differences in impact explained by variations in implementation?
- Much broader - what is the best way to implement a given policy?

5 Common methods for the different aspects of impact evaluation

Different methods are needed for the different elements of an impact evaluation:

- (1) Clarifying objectives and values;
- (2) Developing a theory of change;
- (3) Answering descriptive questions;
- (4) Answering causal questions; and
- (5) Summarising evidence into an overall judgement.

The first of these needs to be part of evaluability assessment work in initiating an impact evaluation. Having decided on the way forward, Table 1 summarises the other questions and the key methods which are then discussed in turn.

² This last point covers questions related to implementation evaluation, when linked with an impact evaluation.

Table 1: Questions and methods for different impact questions

Purposes	Common impact evaluation questions	Common evaluation methods and approaches
Element 1: Clarifying objectives and values	What are desirable impacts and what are negative impacts? What is a desirable distribution of benefits? What are an appropriate (set of) indicators that can help measure these?	<ul style="list-style-type: none"> • Literature review • Desk review • Appreciative Inquiry • Community surveys • Participatory tools with stakeholders • Most Significant Change
Element 2: Developing a theory of change	What is the theory of change underlying the intervention	<ul style="list-style-type: none"> • Theory of Change • Logframe, a results chain or an outcomes hierarchy.
	How is the theory of change working in practice?	<ul style="list-style-type: none"> • Outcome Mapping • Factual analysis • Implementation/process evaluation • Methods below
Element 3: Answering descriptive questions	What has implementation been like (what activities have been undertaken and what has been the quality of implementation?) What agencies, people and mechanisms have been involved in the implementation (or absent in the case of implementation failure) What changes have occurred (and for whom?) What has been the context in which the programme has been implemented?	<ul style="list-style-type: none"> • Re-analysis of existing statistical data • Surveys, administrative data, census data • Observation • Interviews/group interviews/focus groups • Participatory tools • Monitoring data • Process evaluations
Element 4: Answering causal questions	How far has the intervention caused the impacts, contributed to causing the impacts, or have the impacts in fact been caused by other factors? How <i>much</i> of the impact can be attributed to the intervention?	<ul style="list-style-type: none"> • Counterfactual methods <ul style="list-style-type: none"> ○ Randomised control trials ○ Comparison group analysis ○ Regression analyses ○ Logically created or expert constructed counterfactuals • Checking results match tightly the theory that the programme produced them • Identifying and ruling out alternative explanations
Element 5: Summarising evidence into an overall judgement	What is the overall judgement to be drawn from the above data?	<ul style="list-style-type: none"> • Numerical scoring • Rubrics • Cost-effectiveness and cost-utility studies • Consensus consultation/experts' panels

5.1 Clarifying objectives and values (“What would success look like?”)

Impact evaluation draws conclusions about the degree of success (or failure) of an intervention, so it is important to clarify what success looks like in terms of:

- Achieving desirable impacts and avoiding (or at least minimizing) negative impacts;
- Achieving a desirable distribution of benefits.

Formal stated goals and organizational policies are an important start to clarifying what people value (objectives) but by themselves are not sufficient. For example, the stated goals of a road development project might be increased access to markets but it might also produce benefits such as improving access to maternity hospitals. An impact evaluation also needs to take into account any intended and unintended consequences, both positive and negative. What level of loss of biodiversity or negative effect on water quality would be considered a reasonable trade-off for the road? The evaluation should also consider the distribution of the benefits – for example, did the improvements in market access only benefit wealthier farmers who could afford to pay transport costs ie issues of equity, distribution and diversity of impacts?

Different stakeholders may have different views about which values should be used in an evaluation and additional methods will be needed to negotiate amongst them.

Formal documents, such as policies and project plans, can be reviewed to identify the stated values. To identify values which are not formally stated requires additional methods such as Appreciative Inquiry, community surveys, and Most Significant Change. Appreciative Inquiry engages key stakeholders in a discussion around when the programme worked particularly well³. Community surveys would ask community members to nominate or rate the issues that they see as most important in their community. Most Significant Change is a structured process for generating and selecting stories of change that identify what individuals or groups see as the most important results (changes)⁴.

To negotiate between different perspectives about values, different methods will be needed. In a face to face meeting, different forms of voting can be used, including multi-voting with sticky dots, where people allocate their dots against one or more priorities, and dotmocracy, where participants record their level of agreement with each option. The Delphi method uses cycles of written interactions to develop agreement about priorities, Nominal Group Technique and critical incidence analysis are sometimes used to establish consensus (and dissensus) amongst stakeholders. These are just a few examples.

Useful Tip

For more information on how to determine what ‘success’ looks like, check out http://betterevaluation.org/plan/engage_frame/criteria_and_standards

5.2 Developing a Theory of Change

The Theory of Change (TOC) or programme theory describes the causal mechanism of how activities and outputs (such as meals delivered to needy school children) will result in the anticipated outcomes (eg improved concentration in school), and impacts (eg improved grades) and the assumptions involved. There can be multiple TOCs that describe the programme. For example different theories can show how the intervention works in different contexts, or at different stages of the intervention, or even for different intended impacts (Interaction: 6-7). The ToC should be established during the early planning stages of a policy or programme.

A ToC can help to identify which impacts are likely to be achieved during the timeline of an evaluation, and what else should be examined in the evaluation – activities, context, and intermediate outcomes. Also the ToC helps to identify what needs to be in place – people, agencies, activities, mechanisms, resources – for the impact to be achieved. It can also be used

³ For further information go to www.appreciativeinquiry.case.edu

⁴ For further information go to <http://mande.co.uk/special-issues/most-significant-change-msc/>

to analyse the evaluation results. If a programme has not worked, the ToC can help to identify whether this is due to failures in implementation or because the theory of change does not work. If a programme has worked, the ToC can help to identify what is needed to repeat this success at another time or another site.

Developing a TOC is best done through a combination of a desk review of existing documentation, a literature review of research and evaluations of similar programmes including systematic reviews, observing the programme (if it is already running) or similar programmes, and talking with stakeholders about how they think it works. It often involves an iterative, participatory process with programme developers and/or staff and other relevant stakeholders.

The Theory of Change can be represented in the form of a logframe, a results chain or an outcomes hierarchy.

Outcome Mapping is a particular approach to developing a Theory of Change which is particularly suitable when a programme does not directly produce the intended results but works through influencing the behaviour of people in another organisation⁵.

Useful Tip

For more information on different approaches for developing a TOC, check out http://betterevaluation.org/plan/define/develop_logic_model

For evaluations under the National Evaluation Plan, the theory of change should also be expressed in the form of a logframe. Annex 1 of Guideline 2.2.3 on Planning Implementation Programmes provides an example of a theory of change for the National School Nutrition Programme, as well as a model of a logframe. The logframe should include:

- i. Indicators at different levels, baselines and SMART targets, where appropriate, as part of the logical framework;
- ii. The key assumptions and risks which underlie the results chain;
- iii. Key outputs and related activities required to achieve the desired outcomes;
- iv. A summary of the human and financial resources (inputs) needed to achieve the outcomes and impacts.

If a theory of change and logframe does not exist, then one of the first activities in the implementation evaluation should be to derive one, based on the understanding of how the programme or policy was established. As part of the evaluation, changes to the TOC and logframe may be recommended.

5.3 Answering Descriptive Questions

Impact evaluations need to answer descriptive questions – What is the magnitude of the problem (i.e. how many children need to be fed/provided with workbooks, etc.)? Where do the people affected by the policy/programme live/work? How dispersed is the population? What has implementation been like (what activities have been undertaken and what has been the quality of implementation?), what changes have occurred (and for whom?), and what has been the context in which the programme has been implemented? It is likely that a combination of qualitative and quantitative data will be needed to answer these questions well, as part of the implementation evaluation component. Typically, this is administrative data, survey data and

⁵ For further information go to <http://www.outcomemapping.ca/>

census data, all of which can be obtained from Statistics SA, or from HSRC or academic centres of expertise.

5.4 Answering causal questions

In addition to describing what changes have occurred, an impact evaluation must explain whether the intervention contributed to producing these observed impacts. It is rare that a programme or policy is the only cause of identified change. Most policies involve what are called 'complex interventions' that operate at different levels, e.g. the individual, the social/community, the economic, and the macro environment levels. Therefore, 'causal attribution' does not refer to total attribution, but partial attribution or analysing the programme's contribution.

There are three groups of methods for answering causal questions, and an impact evaluation might well use some combination of these:

- Comparing results to a counterfactual – an estimate of what would have happened in the absence of the programme or policy;
- Checking that the results match tightly to the theory of change⁶;
- Identifying and ruling out alternative explanations.

5.4.1 Estimating a counterfactual

Experimental and quasi-experimental research designs estimate the counterfactual by creating a comparison group or a control group – a group who are like the people who received a programme in all ways except for not receiving the programme.

Randomized controlled trials (RCTs)/Control Group

RCTs are an experimental research design which randomises potential participants (or sites) into two or more groups – one (or more) which receive the programme (the 'treatment group') and one which does not (the 'control group'). In theory, after randomisation the two groups should be equivalent, although this should be checked. Sufficiently large samples are required for RCTs to be valid. Data are collected about baselines and final outcomes for the different groups, and these are then compared to show the net impact of the programme. In addition, information is gathered about the quality of implementation of the programme (since failures in implementation might explain a lack of differences between outcomes for the treatment and control groups). The control group might receive no programme, or the current standard programme (often referred to as the 'business-as-usual case'), which is then compared to a new alternative.

RCTs can be appropriate when it is possible to randomise people or sites, where compliance with randomisation procedures can be assured, and where attrition from the experimental and control groups over time can be minimalised. Ethical concerns about withholding the programme from the control group can sometimes be addressed through a design where everyone receives the intervention but rollout is staggered, which may well match government

⁶ Some would see this as not a robust form of Impact Evaluation because it relies on opinion i.e. the eye of the beholder. Similarly, "Checking results match tightly the theory that the programme produced them" and "Identifying and ruling out alternative explanations" will not provide a measure of the benefit.

capacity to rollout. In this case the comparisons are those receiving for longer compared to those receiving for a shorter period.

Comparison group

In quasi-experimental designs, a comparison group is created which attempts to be equivalent to the treatment group, but without using randomisation.

In *matched comparisons*, participants (individuals, organizations or communities) are each matched with a nonparticipant on variables that are thought to be relevant, in an attempt to create equivalent groups without randomisation. Matching can only be done on observable variables with available data (for example, age and gender) so it can be difficult to match on all the important variables (for example, motivation). *Propensity scores* are a particular approach to creating matched comparisons based on an analysis of the factors that influenced people's likelihood (or propensity) to participate in the programme – it is particularly useful when participation is voluntary (for example, watching a television show with health promotion messages).

A *difference-in-difference* design compares the before-after difference for the group receiving the intervention compared to the before-after difference for those who did not.

A *regression-discontinuity design* is useful when there is a threshold for individuals receiving an intervention just below (or above) a threshold. For instance, families may receive a cash benefit, or a service, depending on whether they are above or below a certain level of household income. Families above this threshold can be compared with similar individuals just below the threshold, who are likely to be similar in most other respects and hence comparable

Logically created or expert constructed counterfactuals

Where it is not possible to create a control group or a comparison group, it still might be possible to produce a credible estimate of the counterfactual. In some cases it is credible to use the baseline as a *logically created estimate of the counterfactual*. For example, where a water pump has been installed, it might be reasonable to measure the impact by comparing the time spent getting water from a distant pump before and after the intervention, as there is no credible reason that the time taken would have decreased without the intervention.

Another approach that might be appropriate is to ask key informants to produce an *expert constructed counterfactual* by estimating the likely result if there had been no programme, where they have experience in the usual change trajectory for individuals.

5.4.2 Checking that results match tightly the theory that the programme produced them

Careful checking of the patterns of results against the theory of change can strengthen or weaken the case for causal attribution. This analysis should consider *timing and sequence* – did the impacts occur at a time consistent with the theory of change – not before the intervention was implemented? It should consider *dose-response* - were there better outcomes for participants who received more of the programme (for example, attended more of the workshops or received more support) or better quality services? Did *key informants* (who might include participants) believe the intervention had made a difference, and could they provide a plausible explanation of why this was the case? *Multiple baselines* or *rolling baselines* – where the implementation of an intervention is staggered across time and intervention populations – this can reveal a repeated pattern in each community of a change in the measured outcome

after the intervention is implemented, along with an absence of substantial fluctuations in the data at other time points.

5.4.3 Identifying and ruling out alternative explanations

The third set of methods for causal inference involves identifying and ruling out alternative explanations.

Key informants can be useful to identify possible alternative explanations and these can then be tested using specific data to investigate whether these might plausibly be the explanation. For example a decline in head injuries after the introduction of a bicycle helmet law might indicate it has been successful – or be due to decreased cycling rates. If the number of non-head injuries has remained stable during this time, this would rule out that as an explanation. These methods are generally used, and appropriate, when and where a strong counterfactual is not obtainable.

Tip: Methods for causal attribution

Betterevaluation.org provides information on many different methods, grouped in terms of these three different strategies:

- Estimating a counterfactual
http://betterevaluation.org/plan/understandcauses/compare_results_to_counterfactual
- Detailed comparison to what would be expected if the programme was producing the results
http://betterevaluation.org/plan/understandcauses/check_results_match_theory
- Identifying and ruling out alternative explanations
http://betterevaluation.org/plan/understandcauses/investigate_alternative_explanations

5.5 Synthesis

Finally impact evaluations synthesize information around the specific evaluation and provide an overall evaluative judgement. For example, if a programme has been effective for some but not all participants, or if it has produced positive outcomes but also negative outcomes, these varying results need to be weighed in some way to produce an overall result.

Numeric weighting gives a score for each criterion and then adds them up – usually with some criteria weighted more than others. This is the method usually used to synthesise evidence for evaluating a proposal or a potential employee, but can also be used to evaluate a project, especially to compare projects. This requires some empirical basis for allocating weights, something that can be difficult to establish.

Rubrics produce a detailed scale with ratings using descriptions for each level of performance, which can be linked to a score. Because the criteria are public, a scoring rubric allows stakeholders to all evaluate the criteria, which can be complex and subjective, and they provide a basis for self-evaluation, reflection, and peer review. It is aimed at accurate and fair assessment, fostering understanding and learning. They can be more effective in engaging different people in discussing and negotiating agreement about final evaluative judgements.⁷

Resource

For particular methods for synthesising evidence, see

http://betterevaluation.org/plan/synthesize_value/synthesize_data_single_evaluation

⁷ The South African system of evaluative competences is an example of a rubric.

6 Evaluation Process

6.1 Who undertakes the evaluation

Impact evaluations need to be undertaken by an independent service provider who specialises in research and evaluation to ensure credibility.

6.2 How long should the impact evaluation take

An impact evaluation combined with implementation (recommended) will take a minimum of 6 months, and could take 2-3 years if it includes a baseline and a repeat baseline. The Grade R impact evaluation took around 3-4 months as it only did work on existing data, and had no implementation component.

6.3 Result of the impact evaluation

The process of the impact evaluation is very important as it builds understanding amongst stakeholders on emerging results and commitment to changes to the intervention.

At the end of the evaluation some decisions are needed. These are:

1. Is the intervention achieving its outcomes/impacts and is it value-for-money – if not what needs to be done?
2. How should implementation be strengthened to maximize the likelihood of impact? Should the intervention be upscaled (or downscaled) and what are the implications?

7. Critical issues when planning and managing impact evaluations

This section covers particular challenges that may be encountered in relation to impact evaluations. It draws on the quality criteria set out in the NEPF, specifically: relevance and timeliness, legitimacy, credibility, ethics and trade-offs.

7.1 Relevance and timeliness

Planning for an impact evaluation, and collecting data for an impact evaluation, should be initiated from the beginning of the programme. Impact evaluations should be conducted if their findings will be relevant to future planning, and in time to incorporate the findings into decision-making. In practice when departments want to undertake impact evaluations and this has not planned in advance the data may not be available.

If a programme manager has limited evaluation resources and needs to choose between implementation evaluation and impact or economic evaluation, there may be reasons for choosing implementation evaluation. For example, unless one knows that the programme is being implemented according to design, there may be little reason to expect it to produce the desired outcomes. Results identified without understanding how they were achieved is of very little management use to a programme manager. In some cases there are obvious reasons preventing impact and it is not worth the investment in an impact evaluation. However a note of caution, you may get a well implemented programme that has no positive impact, and may indeed do harm.

For this reason in many cases under the National Evaluation Plan, where an impact evaluation has been requested, in practice it has proved more appropriate to do an implementation evaluation first, and then plan thoroughly to do an impact evaluation at some point in the future.

7.2 Legitimacy

The legitimacy of an impact evaluation can be improved by ensuring that it considers the perspectives of different stakeholders in terms of what would be considered as successful implementation. This might include involving key stakeholders in the development of evaluation questions and the evaluation design, or involving the programme management team in interpreting observation and interview data. This can include beneficiaries, e.g. involving them in the process of sharing their experiences of service delivery through interviews or surveys, or involving them in the process of collecting data, through community score cards, or participatory mapping processes, or the methods of Appreciative Inquiry and Most Significant Change outlined earlier.

Legitimacy comes from explicit and transparent criteria of data extraction and analysis against explicit criteria of internal and external validity, and adequacy of reporting.

7.3 Credibility of the evidence

As in all evaluations, impact evaluations should be explicit about the methods chosen, the reasons for their use, their limitations and how these have been addressed. Key issues to address in terms of credibility are: the quality of existing data; the quality of additional data collected; and sampling. There is also an issue of design bias with some people believing that only RCTs, or some other evaluation design, are able to provide credible evidence of impact. In practice it is often difficult to undertake RCTs for many complex policy issues, and other methodologies are needed, which must still be carried out with rigour. More generally, all evaluation designs carry a risk of bias. Consequently, all evaluation reports should include a risk of bias assessment, and an indication of the degree to which this risk was, or was not, overcome.

Impact evaluations need to assess the quality of existing data used, such as programme reports, media reports, existing photographs and performance indicators. The methods for collecting primary data need to be carefully chosen and implemented appropriately. In particular the expertise and independence of those collecting data needs to be assessed. It is important to check whether data have been collected, and sometimes verified, by an independent agency. Additional data collection should be supported by a combination of expert knowledge about the programme and well-planned and carefully documented data collection, interpretation and analysis.

Data sources for impact evaluations should be chosen so that they triangulate important issues and balance out the limitations of any one source. Sampling decisions should be transparent, and the sampling of informants, sites and time periods should be carefully done to ensure adequate coverage, and any limitations carefully noted.

7.4 Trade-offs

There can be critical trade-offs for different types of impact evaluation designs. A longer intensive design that collects data from all sites may provide answers to every single evaluation question yet it may have high costs. A short, internal evaluation may be cost effective and

provide answers to all the posed evaluation questions, yet lack credibility because it did not have an external evaluator.

8 Typical costs

This section provides some basic guides for helping to determine the size of an evaluation budget. However each context will be unique and require specific budgeting discussions and decisions.

The programme manager has a key role in ensuring that the scope of what is promised by evaluators, or expected by the programme manager, is realistic for the amount budgeted; as over ambitious and under budgeted scope of work is likely to yield a weak base of evidence and an unused report.

Budgeting for an evaluation is dependent on numerous factors. A general 'rule of thumb' is that an evaluation should be between 0.1% to 5% of an intervention's budget. However this depends on many variables such as the amount of credible data already collected, the timeline to collect data, the amount of field work that needs to be done, and other contributing cost factors. IE has a lot of fixed costs. Cost drivers include geographic scope, length of questionnaire, number of respondents, etc and fixed costs (whether done internally or hired out) include preparation of concept notes, ToRs. questionnaires, oversight, etc. while variable costs relate to the scope of the data collection.

Signed



Dr Sean Phillips
Director-General

The Presidency: Department of Performance Monitoring and Evaluation

Date: 31 March 2014

Annex 1: Glossary

Attribution:	A concept in social psychology addressing the processes by which individuals explain the causes of behavior and events.
Comparison Group:	A group of units (e.g., persons, classrooms) that receive either no treatment or an alternative treatment. The purpose of a comparison group is to serve as a source of counterfactual causal inference
Counterfactual:	measures what would have happened to beneficiaries in the absence of the intervention, and impact is estimated by comparing counterfactual outcomes to those observed under the intervention.
Credibility:	The quality of being trusted and believed in.
Impact Evaluation:	Type of evaluation that seeks to measure changes in outcomes (and the well-being of the target population) that are attributable to a specific intervention. Its purpose is to inform high-level officials on the extent to which an intervention should be continued or not, and if there are any potential modifications needed.
Intervention:	The action or process of intervening, for example, a high degree of state intervention in the economy through a programme, policy or plan.
Key Informants:	Those whose social positions in a research setting give them specialist knowledge about other people, processes or happenings that is more extensive, detailed or privileged than ordinary people, and who are therefore particularly valuable sources of information to a researcher, not least in the early stages of a project.
Randomized Controlled Trials:	are specific types of scientific experiments, and the gold standard for a clinical trial. RCTs are often used to test the efficacy or effectiveness of various types of medical interventions within a patient population.
Relevance:	The extent to which the objective of an operation are consistent with beneficiaries 'needs, country needs, organizational priorities and partners and donor policies.
Result Based Management:	A management strategy focusing on performance and achievement of outputs, outcomes and impacts.
Theory of Change:	An explicit presentation of the assumption about how changes are expected to happen within any particular context and in relation to particular intervention.
Trade-offs:	A situation that involves losing one quality or aspect of something in return for gaining another quality or asp

Annex 2: Useful web resources

Asian Development Bank <http://www.adb.org/data/economic-research-initiatives/implementing-impact-evaluation-at-adb>

Better Evaluation: http://betterevaluation.org/evaluation-options/report_friendly_writing

Impact Evaluation in Practice

(http://siteresources.worldbank.org/EXTHDOFFICE/Resources/5485726-1295455628620/Impact_Evaluation_in_Practice.pdf).

Annex 3: Bibliography

- Babbie, E. (2013). *The practice of social research (13th ed.)*. United Kingdom: Wadsworth Cengage Learning.
- Babbie, E., & Mouton, J. (2001). *The practice of social research*. Cape Town: Oxford University Press.
- Bamberger, M, Rugh, J, Mawbry, L, (2012): “*RealWorld Evaluation: Working Under Budget, Time, Data, and Political Constraints*”, SAGE, 2012, 236
- Creswell, J. W., & Plano Clark, V. L. (2011). *Designing and Conducting Mixed Methods Research (2nd ed.)*. Thousand Oaks: SAGE.
- Henning, E., Van Rensburg, W., & Smit, B. (2004). *Finding your way in qualitative research*. Pretoria: Van Schaik.
- Krueger, R. A., & Casey, M. A. (2000). *Focus groups: A practical guide for applied research*. Thousand Oaks: Sage.
- Neuman, W. L. (2006). *Social Research Methods. Qualitative and Quantitative Approaches*. Toronto: Pearson.
- Patton, M. Q. (2002). *Qualitative Research & Evaluation Methods (3rd Ed.)*. Thousand Oaks: SAGE.
- Patton, M.Q. (2008): “*Utilization Focused Evaluation*”, Fourth Edition, Thousand Oaks, SAGE.
- Rogers, P. (2012) “*Introduction to Impact Evaluation*”. Interaction.
<http://www.interaction.org/impact>