Evaluating Security and Justice:

Frequently Asked Questions

Jim Parsons
Aminou Yaya
Caitlin Gokey
Monica Thornton

Date: April 18, 2013

Practice Products for the CCVRI

Improving Measurement in DFID Crime, Conflict & Violence Programming

This document is one of a series of Practice Products developed under the *Conflict, Crime, and Violence Results Initiative (CCVRI)*. The full set of products is intended to support DFID country offices and their partners to develop better measures of programme results in difficult conflict and fragile environments.

DFID recognises the need to focus on the *results* of its work in developing countries. To this end, DFID strives to account better for our efforts on behalf of UK taxpayers, offering clarity regarding the value and impact of our work. The Results Initiative operates under the assumption that we will achieve our development objectives with our national partners more effectively if we generate—collectively—a clear picture of the progress being made.

Within DFID, the Conflict Humanitarian and Security Department has established a partnership with a consortium of leading organisations in the fields of conflict, security and justice to develop more effective approaches to the use of data in the design, implementation and evaluation of programmes that contribute to reducing conflict, crime and violence.

In addition to producing these Practice Products, the consortium has established a Help Desk function to provide direct and customized support to country offices as they endeavour to improve measurement of results in local contexts.

The Help Desk can be accessed by contacting helpdesk@smallarmssurvey.org.

The views expressed in this Practice Product are the sole opinions of the authors and do not necessarily reflect the opinions of all consortia partners. This Practice Product does not reflect an official DFID position.

Members of the consortium

Document Summary

Title:
Evaluating Security and Justice: Frequently Asked Questions

Purpose and intended use of this document:
A starting point for non-researchers seeking simple, jargon-free advice on what evaluations are, how they can be used, how to interpret their results, and when to seek expert advice.

Key questions this document addresses:
Eleven 'frequently asked questions' related to the design and use of evaluations and references to more detailed guidance provided by DFID and other organizations, where relevant.

Key messages/essential "take aways":
The "why, when, and how" of conducting evaluations.

Intended audience of this document (including assumed skill level):
DFID program staff.

Key topics/tags:
Evaluations, process evaluation, outcome evaluation, impact evaluation, evaluations in fragile and conflict-affected settings, security and justice evaluations.

Authors and their organizations:
Jim Parsons, Aminou Yaya, Caitlin Gokey, Monica Thornton - Vera Institute of Justice.

Cross-references to other documents in the series:
Indicators of Activities, Outputs, Outcomes and Impacts in Security and Justice Programming.

Table of Contents

Glossary of Terms

Activities: the actions of DFID staff and their partners that are designed to meet a project's objectives.

Evidenced based practices: approaches to providing interventions or services that have been demonstrated to be effective using evaluation research methods.

Experimental design: a set of methods designed to assess the effect of a project by comparing outcomes for those who receive a service or intervention with a randomly-assigned control group.

Fragile and conflict-affected settings: countries, regions or localities that are either experiencing violent conflict, at risk of conflict, or facing the aftermath of war or other forms of violent upheaval.

Impacts: higher level strategic goals, such as increased access to justice or improvements in public safety.

Indicator: a quantitative or qualitative factor or variable that provides a simple and reliable means to measure achievement, to reflect the changes connected to an intervention, or to help assess the performance of a development actor

Inputs: the raw materials that provide a basis for security and justice programs. Inputs can include money, technical expertise, relationships and personnel.

Objectivity: an important feature of research that is based on the facts and includes steps to reduce the influence of the researcher collecting the data or conducting the analysis on the findings.

Outcome evaluations: the application of research methods to measure the effects of a project or intervention.

Outcomes: the benefits that a project or intervention is designed to deliver.

Outputs: the tangible and intangible products that result from project activities.

Process evaluations: the application of research methods to measure the interventions provided by a project, the populations contacted and other aspects of service delivery.

Quasi-experimental design: a set of methods designed to approximate experimental design studies, used in settings where it is not possible to randomly allocate cases for practical, financial or ethical reasons.

Results chain: a graphical representation of the hypothesized relationship between project inputs, activities, outputs, outcomes and impacts.

Systematic: evaluation activities that follow a pre-defined research plan and adopt standards to ensure objectivity and rigour.

Theory of change: a set of assumptions about the relationship between project activities and goals.

Introduction

Evaluations are no longer the domain solely of researchers with specialized knowledge and training in advanced statistical methods. Donors and aid organizations are increasingly requiring projects to conduct evaluations as a condition of funding, and DFID programme staff are routinely asked to oversee evaluations, work with evaluators, or collect and analyze data themselves. An increasing emphasis on '*evidenced based practices'* means that DFID staff are often called upon to commission and interpret evaluations in order to determine what works in security and justice programming. This focus on evidence reflects global trends in the use of data to ensure that initiatives supported with public funds are held accountable. In the U.K. context, evaluations can help ensure that DFID programs are achieving the aims and objectives laid out in the project business case and that they represent Value for Money.

DFID has issued a number of toolkits, guides and 'how to' notes describing various aspects of conducting evaluations in Fragile and Conflict-Affected Settings (FCAS). This document is not intended to supplant or replace existing guidance. Rather, it is designed to serve as a starting point for non-researchers seeking simple, jargon-free advice on what evaluations tell us, how they can be used, how to interpret their results, and when to seek expert advice. It addresses 13 'frequently asked questions' related to the design and use of evaluations and provides references to more detailed guidance provided by DFID and other organizations, where relevant.

Q1. What is an evaluation?

The Organisation for Economic Co-operation and Development (OECD) Development Assistance Committee defines an evaluation as "the systematic and objective assessment of an on-going or completed project, programme or policy, its design, implementation, and results in relation to specified evaluation criteria."[1] This definition has been widely adopted by DFID and other international development agencies.

As subsequent sections of this document illustrate, evaluations adopt a variety of methods of data collection and analysis; they can be based on quantitative or qualitative data and may use existing data, or information collected specifically for the purpose of evaluating a project. While there is wide diversity in evaluation designs, they are always based on *systematic* data collection, and adopt pre-defined methods of analysis. This sets evaluations apart from other, less structured forms of assessment. For example, while a single visit to a courtroom to observe a legal defense project in action, or a conversation with a DFID project manager about their experiences as part of ongoing disarmament, demobilisation and reintegration (DDR) initiatives, may provide useful information on the successes and challenges of these projects, they would not be considered evaluations. If court observations were organized so that a common set of observers visited selected courtrooms and used pre-defined instruments to collect information, this may provide evaluation data. Similarly, interviews with a sample of DDR project managers that addressed a pre-determined series of questions could be used as part of an evaluation (see Q2). In both of these examples, the key determinant is the fact that data collection was *systematic*; it adhered to a pre-defined set of methods based on a set of research principles designed to ensure the *objectivity* of results. To help ensure objectivity, evaluations are typically conducted by independent organizations that are not involved in providing services and therefore do not have any vested interest in demonstrating that a project or collection of projects is effective.

In addition to these general evaluation principles, when working in FCAS it is important to ensure that any evaluation activities are conflict sensitive – that is they should be based on an understanding of the local conflict situation, designed to minimize potential negative impacts, and maximize the benefits of data collection and reporting activities.[2]

Q2. What can an evaluation tell me?

Evaluations can help answer a range of questions relating to individual projects. For example, you can use evaluation methods to: track the services or products provided by an initiative; assess whether your are reaching your target population; detect unforeseen obstacles that may limit effectiveness; determine the extent to which a project is achieving its intended medium and long-term benefits; and, track the distribution of benefits are across target groups. While evaluations often focus on a single project or intervention, it is also possible to evaluate regional or national initiatives that include multiple projects. While evaluations may provide a final assessment of effectiveness once a project is completed, it is usually important to feed information back to

---

[1] For a more detailed definition of evaluation, see "Principles for evaluation of development assistance", Organisation for Economic Cooperation and Development (OECD) Development Assistance Committee, (Paris, 1991). Available at: http://www.oecd.org/development/evaluation/2755284.pdf .

[2] For a more detailed discussion of conflict sensitive principles, see "Monitoring and evaluating conflict sensitivity: Methodological challenges and practical solutions" Goldwyn (Rachel) & Chigas (Diana). DFID Helpdesk CCVRI consortia, March 2013.

managers and practitioners over the life of an initiative, providing information to maximize impact by adjusting the design and responding to unanticipated challenges.

To get the greatest benefit from your evaluation, the choice of measures and data collection techniques should be tailored to your specific project or projects. Decisions about evaluation methods should therefore start during the design phase with the project theory of change– the hypothesized relationship between your activities and intended outcomes. *Results chains* provide a useful project design tool for mapping out project inputs, activities, outputs, outcome and impacts and can be used to develop evaluation measures (see below for an example results chain). Once you are able to clearly articulate a theory of change and results chain for your project, you will be ready to design evaluation methods that test whether these theoretical relationships hold true in practice – and if not, how the project can be modified or improved.

Figure one: The Results Chain
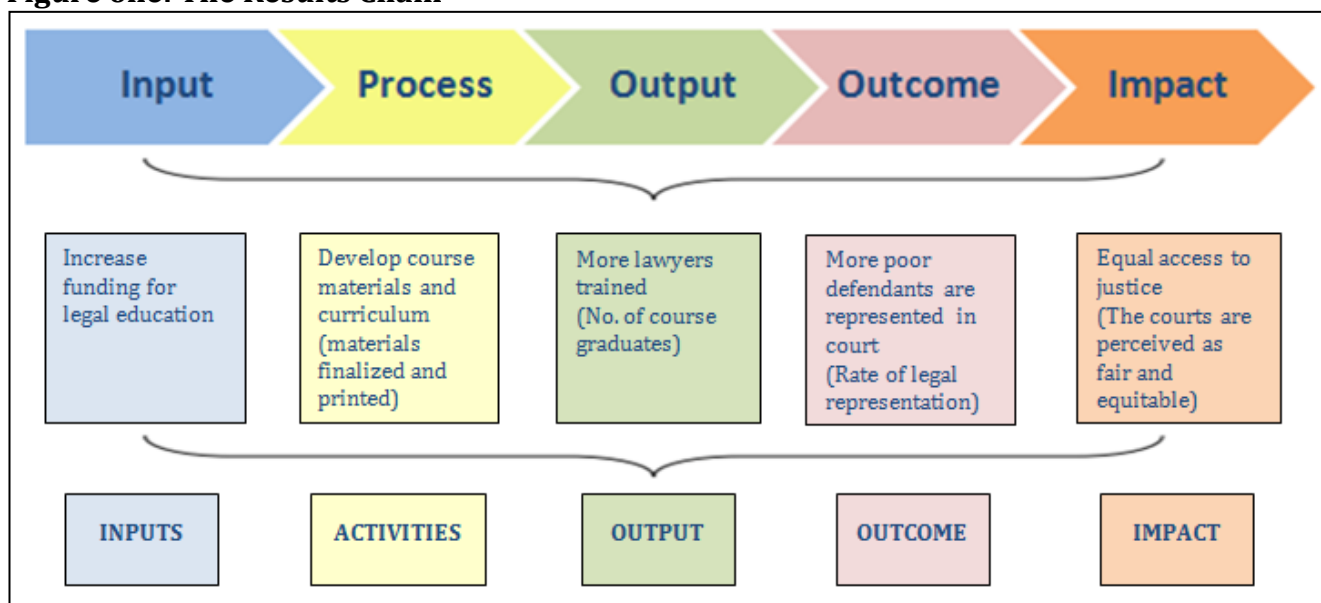
Evaluations can also be used to assess whether a project represents value for money; defined by DFID using the 'three Es' of Economy, Efficiency and Effectiveness. To assess Economy (whether the price paid for labour and other materials represents good value) you will need to collect information on inputs. To evaluate Efficiency (the relationship between cost and outputs) you will need to measure inputs and outputs. To assess project effectiveness (whether an investment is leading to the intended results), you will need to track project inputs, outputs and outcomes.

Q3. What is the difference between a process evaluation and an outcome or impact evaluation?

While the term is often used to refer to a set of methods designed to determine a project's *outcomes* and *impacts* (the degree to which a project is achieving its medium-term and long-term goals), evaluations can also describe the *process* of delivering an intervention or service. These different types of evaluation are complementary and often use similar data collection methods, but their aims are different. In order to track a programs operation along the results chain (from inputs to impacts) you will need to use both process and outcome measures.

Process evaluations provide information on inputs, activities and outputs, including: the individuals and organizations that are involved in setting up and managing the project; services provided; the number and location of project sites; the characteristics of clients served in each site; and whether the project is meeting predetermined milestones or benchmarks. Process data is typically collected throughout the life of a project, providing a valuable feedback loop to allow managers and staff to detect emergent challenges and take corrective action. For example, process evaluation data from a policing project indicating that citizens from a particular racial or ethnic group are reporting crime at lower rates than others might suggest the need to strengthen outreach to underserved communities. Process evaluations use a variety of data sources. They often include information on the number of people served by a project (client flow), collected as part of project management data systems and qualitative interviews with project staff, stakeholders and clients.

Process evaluations and program monitoring: Process evaluations have a number of similarities with program monitoring activities (and often use the same data sources). Both describe the operation of programs rather than their outcomes. However, process evaluation measures are typically designed to generate information that can be used to improve and fine-tune program activities, whereas monitoring measures are primarily intended to ensure accountability and as a way of determining whether projects are "on course". Process evaluations are often combined with outcome measures, to provide a comprehensive view of project operations, whereas monitoring exercise often stand-alone. Trained researchers usually conduct process evaluations whereas monitoring is typically the responsibility of project managers.

As a general rule, it is usually important to include some process measures in an evaluation. Without information on process, evaluators may find that the project succeeds some of the time (as is often the case), but have no information on what distinguishes success from failure. For this reason, evaluations often combine measures of process and outcomes/impacts to describe how the program was implemented, what it achieved, and who benefited. If you are conducting a VFM assessment, you will need measures of program activity (e.g. the resources that your project uses and what it produces) in order to determine if your project meets standards of economy, efficiency or effectiveness.

Outcome evaluations are usually the only way of knowing definitively whether a project is meeting its goals. [3] They often compare outcomes for an intervention group (the clients of a project or places where the project is active) with a group that is similar but has not received any services. For example, to assess the effectiveness of a series of training seminars on police investigative techniques, evaluators may decide to randomly assign potential participants to either attend the seminars or act as a '*control group,*' who receive no training.[4] If the decision to allocate police officers between the two groups is carefully controlled and truly random, differences in the outcome of interest (e.g. successful investigations) between those receiving training and the control group may be attributed to the intervention.[5] However, in most security and justice setting it is not possible to randomly assign people to receive an intervention, for a range of logistical, political and ethical reasons. There are other methods, referred to as '*quasi-experimental*' designs that approximate experimental studies and help determine outcomes and impacts. There is also a growing body of outcome or impact evaluation methods that draw upon qualitative data, including

---

[3] For a more detailed discussion of outcome or impact evaluation methods, see "Results in Fragile and Conflict-Affected States and Situations." DFID How to Note, 28 February 2012. Available at:
https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/67437/managing-results-conflict-affected-fragile-states.pdf.
[4] For a more detailed discussion on the use of control groups in evaluation, see "Broadening the Range of Designs and Methods for Impact Evaluations." DFID Working Paper 38, April 2012. Available at:
https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/67427/design-method-impact-eval.pdf.
[5] The decision about whether observed differences are meaningful depends on a number of factors, including the size of the difference between the two groups and the number of people involved in the evaluation.

'case-based' and participatory approaches. These methods use detailed descriptions of contextual data to examine the conditions that are required in order for a project to succeed. [6] Participatory approaches ask the intended beneficiaries of a project or programme to rate success. This can generate new insights into the strengths and challenges of security and justice programming, while also increasing ownership and buy-in amongst target populations.

Given the cost and logistical challenges associated with experimental and quasi-experimental designs, performance indicators are increasingly used as proxies for project effectiveness (see Q4).

Q4. How do evaluations differ from performance indicator projects?

DFID defines an indicator as "a quantitative or qualitative factor or variable that provides a simple and reliable means to measure achievement, to reflect the changes connected to an intervention, or to help assess the performance of a development actor." Many projects are required to collect indicators to track programme performance. By combining measures of inputs, activities and outputs indicators can be used to monitor the extent to which a project is operating as planned (see Q2 for more information on these terms). Comprehensive sets of indicators can also provide information on the *association* between activities and outcomes or impacts. However, indicators will not tell you if observed changes can be *attributed* to a particular project, or set of projects. *Evaluations are typically designed to provide a detailed description of project delivery and project effectiveness and often address questions of attribution*.

For example, a mediation project may choose to compare the time taken to resolve a dispute via mediation with the equivalent period for cases in the formal court system as a performance indicator. This indicator may provide an overall sense of whether cases handled by the mediation services were resolved more quickly, and could be particularly informative if used to monitor ongoing changes in case duration over time. However, taken in isolation, it would not provide definitive information on whether an observed reduction in case processing could be attributed to project activities. This is because indicators are not designed to control for alternative explanations that may influence findings. For example, the formal courts may choose to refer less complex cases for mediation that would take less time to resolve in either system. It could also be that both parties are required to agree to mediation before a case is referred, further skewing the analysis towards less contentious cases. Or a statutory change that guarantees legal representation in the formal court system could delay cases, but improve outcomes. *Evaluation methods are designed to account for these competing explanations and can provide definitive answers to questions about whether a project achieved pre-defined goals and in which cases it worked*. A good evaluation can also provide information on reasons for failure, if a project is unsuccessful or only partially successful. This kind of information is particularly valuable when designing new projects.

Q5. What kind of data can I use to evaluate my project or programme?

There is a range of data sources that you can draw upon when conducting an evaluation, depending on the focus of your project, the resources that you have available, and the type of information required (e.g. process, outcomes or impacts). Broadly speaking, data sources can be divided

---

[6] For a more detailed discussion of the range of approaches to outcome and impact evaluation, see "Broadening the Range of Designs and Methods for Impact Evaluations." DFID Working Paper 38, April 2012. Available at: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/67427/design-method-impact-eval.pdf.

between those that generate quantitative and qualitative information. These methods should not be seen as mutually exclusive and evaluations often include quantitative and qualitative data. Combining data sources in this way can provide checks and balances on the biases that can result from reliance on a single type of data. The first step in selecting data sources is to clearly articulate the aims and objectives of your project, and the conclusions that you hope to draw from your analysis.

Quantitative data is used to provide a numerical summary of processes, outcomes or impacts related to your project (e.g. quantity, incidence, duration or frequency). For example, the average period of pretrial detention, the budget of a national prison system, the time that it takes for the police to respond to an emergency call for service or the proportion of judges that are female are all quantitative measures. Evaluators often use quantitative data because it can be analyzed using methods that provide a statistical estimate of project effectiveness (i.e. whether the project was associated with a measurable improvement). However, the process of quantification can also mask important nuances that may be important to describe as part of the evaluation. Consider, for example, the difference between a report based on detailed face-to-face interviews with female crime victims about their interactions with the police (a qualitative measure) and a single percentage describing the proportion of women responding to a public survey who felt that they were "treated fairly by the police" (a quantitative measure).

Quantitative data sources often include: fixed response or multiple-choice survey questions; administrative records; budgets, strategic plans and other documents; counts or other types of numerical information gathered as part of field observations.

Qualitative data provide a more detailed description of experiences, views, perceptions and situations that are not amenable to quantification. For example, qualitative data could be used to describe the views of project clients about the effectiveness of a service, document the content of discussions between judges and defendants in court, or generate detailed descriptions based on observations of interactions between the police and community members. Qualitative data may be based on hand written notes or recorded conversations and, if the latter, can be used to describe people's views or experiences in their own words. This type of data collection is intensive and even a few interviews or observations can create a great deal of information that requires significant time to process and analyze. Whereas quantitative studies are usually designed to be representative, and therefore require large samples, qualitative research does not claim to be generalizable. Therefore, most qualitative analysis provides a detailed description of a smaller number of cases. While the exact number varies, many qualitative studies are based on 20 or 30 in-depth observations or interviews. Because of its detailed nature, qualitative data is often used as part of process evaluations, to document the experiences of clients, project staff members, government partners and other key project stakeholders, and as a way of gathering suggestions for project improvement.

Qualitative data sources include: open-ended questions from in-depth interviews and group discussions (focus groups); narrative reports based on newspaper articles, monitoring reports and other written documents; narrative analysis of court transcripts; and field observation notes.

Q6. Can I use existing data to evaluate my project?

In many settings you will be able to find existing data that you can use to design and evaluate your project. For example, if you are evaluating a court improvement project you may be able to find information in court records, describing the number and type of cases, characteristics of defendants, availability of lawyers and/or case outcomes. An important first step when thinking

about evaluation is to talk to your partners and colleagues in national government, civil society, and other development agencies about the information that they hold. Even if the data that you need to evaluate your project does not currently exist, you may be able to supplement existing activities to collect the information you need. For example, you may be able to add a small number of questions about rates of victimization to an ongoing national public survey, or ask your DFID colleagues to collect information that is relevant to your project as part of their routine monitoring and oversight activities (e.g. collecting data on a corrections project while conducting routine prison visits). Existing data is particularly useful as a baseline measure, defining the conditions that you hope your project will improve upon.

All data sources have their limitations, and a lack of capacity to collect and store information may mean that the data that you receive from national and international partners is incomplete or inaccurate. This may be a particular challenge in countries without a history of data collection and when working in FCAS. Collecting information from multiple sources can help identify data limitations and, in some cases, may allow you to correct mistakes. For example, you may choose to conduct observations in courtrooms or interview defence attorneys with experience in the courts as a check on administrative records describing case outcomes. You could also compare the same type of information for different sources, such as information from the national law enforcement agency describing the number of complaints filed against the police provided and similar records from an independent ombudsman or civil society organization.[7] Because these agencies have different incentives to report accurate information, comparing information from these sources might detect potential under-reporting (or over-reporting) of the number of complaints.

You will need to collect information on your own project as part of the process evaluation including project expenditure, number of people served, location of project sites, etc. If possible, this information should be included as part of existing management systems to minimize the data collection burden on project staff.

Q7. Should I conduct an outcome evaluation?

Evaluating a project typically requires a significant investment of time and money, and may divert resources that could otherwise be used to support new initiatives or expand existing projects. However, without collecting evaluation data you will typically have no way of knowing if your investments in a new project are delivering any of the anticipated benefits. If you are designing a small project that does not initially warrant the level of investment required to conduct a detailed outcome evaluation, you may initially decide to rely on *process evaluation* measures - in addition to any ongoing program monitoring activities. By documenting the services provided, clients served and obstacles encountered by a new project, process evaluations can provide some of the information necessary to replicate and expand a project while also laying the foundation for subsequent outcome evaluations.

There are several factors to consider when deciding whether to invest in an outcome evaluation, including the size of the project, the resources that you have available, the extent to which the

---

[7] For a more detailed discussion of methods for using data from multiple sources as a check on data problems, see "Guidance on Evaluation and Review for DFID Staff." DFID Evaluation Department, July 2005, p. 79-80. Available at: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/67851/guidance-evaluation.pdf.

project is innovative or tests a new approach, and the ease and cost of evaluation (see boxed text, below). [8]

<div style="border:1px solid">

Deciding When to Invest in an Outcome Evaluation

- The interest among local stakeholders, or where local stakeholders have a clear preference for an evaluation
- The strength of the evidence base for the particular development context
- Whether the programme is innovative, or designed as a pilot, with plans to scale up or replicate
- Any specific cross-cutting concerns, for example anti-corruption or value for money
- Level of contentiousness and risk (there may be views about the applicability of the programme to a particular context)
- Opportunities to work with partners, including supporting capacity, building in evaluation or where programmes are being undertaken jointly
- Financial value (large programmes should be considered for accountability reasons, even if they do not meet other evaluation criteria)
- Whether the programme has particular strategic importance to UK objectives or is a particular policy priority at the country level
- Evaluability (whether it's possible to realistically evaluate a programme)

Source: *Learning What Works to Improve Lives, The UK Government's Policy for Evaluation of International Development*

</div>

Q8. When do I need to consult with an evaluation specialist?

Depending on your level of familiarity and comfort with evaluation methods and the skills within your team, you may need to seek external advice from an evaluation specialist. It is usually important to seek expert advice at the design phase, to ensure that your choice of evaluation measures and data collection methods are attuned to the needs of your project. If you are planning to use a survey as a measure of general public opinion you will need to conduct a 'power analysis' to determine the number of people that you will need to survey in order draw valid conclusions.[9] Similarly, designing methods for selecting people to participate in surveys (sampling) is usually technical and may require expert input.[10]

Ideally, your consultants will be locally based with experience working on issues related to your project. Local universities and civil society organizations are often good places to find local researchers who understand the context where you are working and are able to provide independent, unbiased advice. While this may not always be possible, at a minimum it is important to identify researchers with experience working in FCAS to advise you on design and analysis issues. If your technical advisors are not familiar with the country where you are working, you should also find informal consultants with local experience to provide feedback on the feasibility

---

[8] For a more detailed discussion of deciding which programmes to evaluate, see "Learning What Works to Improve Lives, The UK Government's Policy for Evaluation of International Development." Draft for Consultation, August 2012. Available at: http://mande.co.uk/2012/uncategorized/draft-dfid-evaluation-policy-learning-what-works-to-improve-lives/.

[9] Power analysis is a statistical technique that uses information on the target population to estimate the sample size required to draw conclusions with a specified level of accuracy. For further details, see http://psych.wisc.edu/henriques/power.html

[10] Sampling describes a set of methods designed to ensure that a group of people or cases selected for inclusion in a study are either a) representative of the general population, or b) include important subgroups. Common sampling techniques include probability (random) sampling, cluster sampling, stratified sampling and purposive sampling. For further details, see http://www.socialresearchmethods.net/kb/sampling.php

and cultural appropriateness of your research design and data collection methods – this is particularly important when developing questionnaires, consent forms and other written materials.

Q9. When should I start evaluating my project?

As already noted, it is important to start thinking about evaluation from the beginning of the project design phase and as you develop your theory of change (see Q2). It is important to consider two basic, evaluation related questions: 1) what process measures will you need to collect throughout the life of the project in order to detect and overcome potential obstacles to implementation?; and 2) what are the impact and outcome measures that will indicate whether the project is successful? The answer to the first questions will help structure your initial data collection priorities. Interviews with project staff and clients often provide valuable information on early implementation challenges, generating information to fine tune and improve your project.

Deciding when to start collecting data for an outcome or impact evaluation is a balancing act. On one hand, security and justice projects operating in FCAS typically adapt and change to overcome unexpected obstacles during the first few months of operation and it is usually advisable to allow time for a new project to 'bed-in' before conducting a large-scale evaluation. On the other, funders and project partners may be impatient for information on results in order to make funding decisions and set programmatic priorities. In some cases, outcomes and impacts may take time to emerge. For example, it may take several years before a project that provides vocational training to former combatants delivers measurable reductions in rates of violent crime. In this case, you may be able to find a *proxy*; a measure that can be expected to correlate with the ultimate outcome. For example, you may choose to measure the number of former combatants finding employment as a result of vocational training or changes in literacy rates as a proxy for reductions in violence.

Q10. How much should I budget to evaluate my project?

There are no hard and fast rules about the amount that an evaluation will cost. DFID guidance recommends that evaluation budgets are in the range of 1-5% of total program spend, but also acknowledge that it is "important not to be formulaic in evaluation budgeting, comparing the evaluation cost against the programme cost, but rather considering the value of the knowledge it will yield."[11] Following this guidance, you may choose to invest more of your budget in evaluations of innovative programmes and those that are candidates for expansion.[12]

The cost of evaluations can vary widely depending on a range of factors related to the design and scale of your project, the number of jurisdictions or states where you are operating, the role and affiliation of the evaluator, and the type of data that they will collect. As with most areas of project design, it is important to tailor the scope and goals of your evaluation activities to available funds. Including comparison or control groups (experimental or quasi-experimental designs) and large public surveys can be particularly expensive, and you should ensure that you have sufficient

---

[11] For a more detailed discussion on budgeting for evaluation, see "Results in Fragile and Conflict-Affected States and Situations." DFID How to Note, 28 February 2012 Available at: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/67437/managing-results-conflict-affected-fragile-states.pdf.

[12] For a more detailed discussion on the budgetary considerations of evaluation, see "Learning What Works to Improve Lives, The UK Government's Policy for Evaluation of International Development." Draft for Consultation, August 2012.

resources before embarking on these designs.[13] Working with international evaluation consultants and statistical experts is also generally more expensive than relying on local capacity and you may be able to produce high quality results while minimizing cost by working with government departments, academic institutions or civil society organizations (see question 8 for further details on the advantages of working with local researchers).

Common evaluation budget items

- Costs related to the design of questionnaires and other study instruments
- Travel costs (both international and within the country where you are working)
- The daily rate, travel and subsistence cost of your evaluator (if you are contracting the work out)
- Data collection and data entry
- Translation of questionnaires, consent forms and other study materials
- Costs associated with printing and disseminating your final report

---

Q11. What steps should I take to ensure the safety and wellbeing of research participants?

Participation in a research study can be harmful for your study respondents and evaluators have additional responsibilities to safeguard research participants against potential harm, above and beyond the international human rights principles that govern all development work. For example, asking a government official about a sensitive political issue as part of an expert interview may damage her/his reputation or place them at risk of violent retaliation if the information is not treated confidentially. Alternatively, interviewing women about their experience of domestic violence, may exacerbate the harm caused by the initial victimization if the researchers are not trained in appropriate methods for collecting information on sensitive topics. Qualitative interviews, where respondents are asked to describe their experiences in detail, may be particularly traumatic for survivors of violent crime, if the research is not conducted in a responsible and sensitive manner.

DFIDs *Ethics Principles for Research and Evaluation* lists ten principles for ensuring the ethical conduct of research and minimizing risk.[14] These include the importance of protecting the anonymity of research participants, ensuring that the research is "relevant and high quality," and adhering to any requirements for ethical review (such as approval of research instruments by a local Institutional Review Board or Human Subjects Research Committee) before commencing data collection. When interviewing members of the public, it is generally good practice to provide a written summary or "consent form" detailing the purpose of the research, the fact that participation is voluntary, and any potential risks and benefits associated with participation. You should ensure the confidentiality of research data by only recording people's names or other personal identifiers

---

[13] For a more detailed summary of common evaluation costs see "Designing For Results: Integrating Monitoring And Evaluation In Conflict Transformation Programs" Chapter 8 "Evaluation Preparation." p. 133. Church (Cheyanne) and Rogers (Mark M.), Search For Common Ground 2006. Available at: http://www.sfcg.org/Documents/manualpart1.pdf .

[14] For a more detailed guidance on ethical considerations in evaluation, see "DFID Ethics Principles for Research and Evaluation." DFID July 2011. Available at:
https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/67483/dfid-ethics-prcpls-rsrch-eval.pdf.

if absolutely necessary, storing information securely, and destroying identifiers at the conclusion of the evaluation.

The DFID ethics principles emphasize the importance of ensuring participation of women and other vulnerable groups.[15] However, in some settings, female research participants may be at heightened risk. For example, the mere fact that a women in an abusive relationship has participated in a study on domestic violence may place them at heightened risk of further abuse. When conducting evaluations of this sort, additional steps may be required to ensure the safety of participants, such as ensuring that women are interviewed by female interviewers, arranging interviews in private settings, and identifying the research as a study of 'women's health' or a similarly non-sensitive issue.[16]

Q12. How can I identify local capacity to assist with evaluation activities?

There are a number of reasons to include local researchers and fieldworkers in your evaluation team. Working with local organizations in FCAS keeps investments in evaluation within the country where you are working and can provide an opportunity for researchers to develop their skills and experience.[17] Local evaluators can also provide a more cost effective alternative to employing international evaluators. Perhaps most importantly, local researchers bring contextual knowledge about their country which can help ensure that your methods are attuned to local needs and inform interpretation of findings at the analysis stages of an evaluation. Additionally, by working with local fieldwork partners, you can ensure that any translations of questionnaires and other materials used as part of the evaluation are accurate, and that technical terms used in the research make sense to a local lay audience.

In some cases, you will be able to find a local evaluation expert who is qualified to oversee all aspects of the evaluation design, data collection, analysis and reporting. In others, you may need to seek the services of an external research consultant to help with evaluation design and provide training to local research staff. Many governments have a National Statistical Office or similar department responsible for coordinating national data collection activities that can help you identify researchers. Academics in social science or law departments of local universities may also be well placed to manage evaluations, and will often have a good sense of research organizations that are active in the country. In most major cities you will be able to find civil society groups that focus on justice, gender issues, human rights, or specialize in designing and conducting public surveys. One easy way to identify these organizations is by reviewing existing research reports and talking to your partners who have commissioned similar projects in the past.

---

[15] Ibid.

[16] For a more detailed discussion of ethical considerations when researching domestic violence see "Putting Women First: Ethical and Safety Recommendations for Research on Domestic Violence Against Women." WHO 2001. Available at: http://whqlibdoc.who.int/hq/2001/WHO_FCH_GWH_01.1.pdf.

[17] The Paris Declaration emphasizes the importance of developing evaluation capacity within partner countries with the aim of developing independent monitoring and evaluation functions; for more discussion on using local capacity in evaluation, see "Results in Fragile and Conflict-Affected States and Situations." DFID How to Note, 28 February 2012. Available at: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/67437/managing-results-conflict-affected-fragile-states.pdf.

Q13. What information should I include in my evaluation report?

An evaluation report should include enough information for others to be able to understand the evaluation goals, the nature of the information that was collected, and your findings and conclusions. [18] As a general rule, the description of your methods should be sufficiently detailed that another researcher could replicate the study, including details of the questions that you asked, the data sources and (if relevant) your survey-sampling strategy or other relevant methods for selecting participants. It is particularly important to include any potential weaknesses in your evaluation design, such as missing or unreliable data and geographical areas that were inaccessible to data collectors. You may choose to include information on the methods and more detailed findings (such as statistical tables) in a technical appendix to the main report.

If it is important to communicate results to a policy-maker audience, then you may want to produce a short executive summary or policy-brief, detailing the main findings and policy relevant recommendations without including the level of methodological detail included in the full report. As part of your communications plan, you may choose to organize community forums or produce press releases, as a way of communicating findings to members of the general public.

> **A final evaluation report should include:**
>
> - ✓ Executive summary
> - ✓ Acknowledgements
> - ✓ Project description
> - ✓ Evaluation goals
> - ✓ Methods
> - ✓ Any caveats related to your choice of methods
> - ✓ Findings
> - ✓ Recommendations
> - ✓ Technical appendices

---

[18] For a more detailed discussion of evaluation reports, see "Guidance on Evaluation and Review for DFID Staff." DFID Evaluation Department, July 2005. Available at:
https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/67851/guidance-evaluation.pdf.