



## How to monitor and evaluate anti- corruption agencies: Guidelines for agencies, donors, and evaluators

Jesper Johnsen, Hannes Hechler,  
Luís De Sousa and Harald Mathisen (team leader)

U4 is a web-based resource centre for development practitioners who wish to effectively address corruption challenges in their work.

U4 is operated by the Chr. Michelsen Institute (CMI) – an independent centre for research on international development and policy – and is funded by AusAID (Australia), BTC (Belgium), CIDA (Canada), DFID (UK), GIZ (Germany), Norad (Norway), Sida (Sweden) and The Netherlands Ministry of Foreign Affairs.

All views expressed in this Issue are those of the author(s), and do not necessarily reflect the opinions of the U4 Partner Agencies or CMI/U4. (Copyright 2011 - CMI/U4)

# **How to monitor and evaluate anti-corruption agencies:** Guidelines for agencies, donors, and evaluators

By

Jesper Johnson  
Hannes Hechler  
Luís De Sousa  
Harald Mathisen (team leader)

U4 Issue

September 2011 No 8





# Contents

- Acronyms ..... iv**
- Abstract ..... v**
- 1. Introduction..... 1**
  - 1.1 Organisation of the report..... 3
- 2. Current practices in conducting evaluations ..... 5**
  - 2.1 Understanding the terminology ..... 5
  - 2.2 Results of the analysis ..... 10
  - 2.3 Benefits of better evaluation..... 13
- 3. Why M&E matters for ACAs ..... 15**
- 4. Challenges in ACA evaluability ..... 17**
  - 4.1 Understanding the evaluability of ACAs ..... 17
  - 4.2 Increasing the evaluability of ACAs ..... 21
- 5. How can we perform better evaluations? ..... 30**
  - 5.1 Purpose and rationale for evaluation ..... 31
  - 5.2 Composition of the evaluation team ..... 31
  - 5.3 Evaluation scope, overall principles, methodology, and questions..... 32
  - 5.4 Analytical methods for evaluation..... 35
  - 5.5 Participation ..... 36
  - 5.6 Quality assurance ..... 37
  - 5.7 Budget and financial aspects of M&E ..... 37
- 6. Impact evaluation methodology ..... 38**
  - 6.1 Elements of an ACA impact evaluation model ..... 39
  - 6.2 Current practices in measuring ACA impact ..... 42
  - 6.3 Issues of accountability and attribution ..... 43
  - 6.4 The principle of the counterfactual ..... 45
- 7. Indicators to track performance: Introducing the indicator catalogue ..... 47**
  - 7.1 Construction of indicators ..... 47
  - 7.2 Sample indicators used by three ACAs ..... 49
  - 7.3 Guidelines for use of the indicator catalogue ..... 52
- 8. Learning from M&E..... 56**
- Annex 1. Indicator catalogue..... 58**
- References ..... 73**

## Acronyms

ACA	anti-corruption agency
CoC	code of conduct
CPI	Corruption Perceptions Index (of Transparency International)
CPIA	Country Policy and Institutional Assessment (of World Bank)
DAC	Development Assistance Committee (of the OECD)
DFID	UK Department for International Development
ICAC	Independent Commission Against Corruption (Hong Kong)
KNAB	Corruption Prevention and Combating Bureau (Latvia)
M&E	monitoring and evaluation
MIS	management information system
NGO	nongovernmental organisation
OECD	Organisation for Economic Co-operation and Development
SFO	Serious Fraud Office (United Kingdom)
TNA	training needs assessment
ToC	theory of change
ToR	terms of reference
UNDP	United Nations Development Programme
WGI	Worldwide Governance Indicators (of World Bank Institute)

## Acknowledgements

The research and analysis undertaken in this publication received special support and encouragement from the Norwegian Agency For Development Cooperation (Noard).

U4 would like to thank Anette Simonsen (monitoring and evaluation), Cathy Sunshine (writing and editing) and Lisa Arnstad (U4) who helped us make the report user friendly. Phil Mason (DfID), Francesca Recanatini (World Bank), Anga Timilsina (UNDP) and Phil Matsheza (UNDP) all gave valuable contributions and comradely critique. The authors bear sole responsibility for the content of the report.

This report was reviewed and validated at a UNDP sponsored workshop in Bratislava on 7-8 July, 2011. U4 would like to thank UNDP for this opportunity and hope that the organisation can make good use of the report in its ongoing work to support ACAs globally.

In association with:



## Abstract

The number of Anti-corruption agencies (ACAs) around the world has increased dramatically over the past decades. Nevertheless, the value of ACAs is increasingly being questioned by international donors and national governments. Frequently, ACAs are not considered to deliver on the high expectations bestowed upon them.

Evaluations of individual agencies were collected and analysed to assess the evidence underlying the assumptions about the effectiveness of ACAs. Surprisingly, few evaluations had actually been done, and even fewer measured the actual outcomes and impacts of the ACA. Thus, whilst opinions about ACAs are many, the actual evidence about their performance is scarce. To develop this body of evidence, ACAs need to do a better job at establishing results-based indicators for their work, showing how activities lead to impact, and collecting data.

To which extent the perceived failure of ACAs is an issue of measurement or design can therefore not be answered with any certainty. The value of ACAs can only be determined once evidence-based evaluations are conducted.

To this end, the report provides technical, methodological, and practical guidance to assist staff of ACAs in undertaking monitoring and evaluation and shows how the outcomes and impact of the work of ACAs can be evaluated in an objective, evidence-based manner.





# 1. Introduction

This manual provides technical, methodological, and practical guidance to assist staff of anti-corruption agencies (ACAs) in undertaking an internal process of monitoring and evaluation. It informs donors about how the performance of ACAs can be credibly evaluated and offers a methodology that can be used to ensure that the outcomes and impact of the work of ACAs is evaluated in an objective, evidence-based manner. Based on a mapping exercise of existing evaluations of ACAs, the manual provides suggestions for how such evaluations can be improved in the future. It also recognises the essential task of building up the internal monitoring systems and processes within ACAs, showing how this can be done in a cost-effective manner that facilitates production of useful data.

While the principal focus of the manual is on practical advice for its three main audiences—ACAs, donors, and evaluators—it also addresses policy issues particularly relevant for donor agencies. This is necessary because of the infancy of monitoring and evaluation (M&E) in the anti-corruption field. Better monitoring mechanisms and evaluations are needed to advance the fight against corruption. Poorly designed and poorly implemented evaluations are at best useless; at worst, they are used for window dressing. It is better to conduct a few good evaluations than many questionable ones. Measuring performance, and particularly impact, is challenging, but it is not impossible. We therefore see a role for international donors in integrating more advanced M&E processes into their support packages, thus increasing internal capacities of ACAs. Improved M&E systems require medium- to long-term technical and financial assistance from donors.

Our review of several evaluations of ACAs shows that donors are genuinely interested in assessing impact, but it also shows that they rarely commit the resources and time needed to obtain the high-quality evaluation they are demanding. It is often assumed that organisational assessment exercises can evaluate impact, but this is impossible because the methodologies are not designed for impact measurement.

Reports about failed ACAs continue to make headlines. The rhetoric surrounding the performance of these bodies has changed radically from enthusiastic support to defeatism. After a honeymoon of political commitment and an initial launch with fanfare and high expectations, the story of ACAs is now told as one of stalemate, lack of credibility, and marginalisation. In 2005, a United Nations report concluded,

Several countries have opted for or are currently considering creating an independent commission or agency charged with the overall responsibility of combating corruption. However, the creation of such an institution is not a panacea to the scourge of corruption. There are actually very few examples of successful independent anti-corruption commissions/agencies. (UNDP 2005, 5)

Representatives of individual agencies, not surprisingly, routinely challenge the validity of negative statements about the performance and impact of ACAs. Pundits debate whether the perceived failure of some ACAs is a measurement issue or a design issue, as measurement issues can be overcome more easily than design issues. It is safe to say that at this point, no one knows. An informed debate about the value of ACAs can only begin when credible, evidence-based evaluations, derived from improved M&E, are available.

In comparison to other sectors of government, monitoring and evaluation of anti-corruption activities is generally neglected. Introducing basic processes and systems will require some investment in human resources, but the costs of improving M&E significantly, even if not sufficient for a comprehensive assessment of impact, should not be insurmountable. The principal requirement is a change in mind-set.

ACAs are never given an easy or a well-defined task. They are expected to combat corruption in an independent, knowledge-based manner by developing specialized enforcement competences along with preventive and educational/research capacity. They have to overcome the inadequacy of traditional law enforcement structures and processes and assume a leading role in implementing national anti-corruption strategies. Last but not least, they must reassure the public of the government's commitment to fighting corruption (Meagher 2004).

ACAs may be set up in an attempt to upgrade a country's ethics infrastructure, respond to corruption scandals, or fulfil international treaty requirements (Meagher 2002; De Sousa 2009a, 2009b). Launch of an ACA may even be a political manoeuvre by populist executives to control the anti-corruption discourse (Smilov 2009). Some ACAs are set up under donor and international pressure in a context of state failure or state-building programs. Critics point out that expecting ACAs to deliver when other governance structures, including supporting structures, are nonexistent, underdeveloped, or underperforming is "mission impossible," a challenge that no agency could be expected to meet (Doig 2009). Measuring the performance and impact of ACAs, therefore, requires careful attention to what they can be expected to achieve and what specific outcomes are feasible.

The mapping exercise of ACA evaluations showed that very few evaluations have been conducted, and that the quality of these evaluations is generally very low. This is not a new problem, but rather one that has been well known for some time. In 2005, after reviewing five African ACAs, Doig, Watt, and Williams (2005, 5) summarized the situation as follows:

A central problem is the measurement of ACC [anti-corruption commission] performance—in particular the lack of appropriate measurement tools and the widespread employment of inappropriate, unhelpful, unrealistic and even counter-productive measures of performance. This creates a further difficulty in differentiating between achievable and non-achievable organisational performance and compounds the problem of distinguishing between factors which are within the ACC's control and those that are not. Donors do not actually know if their funding has any impact on corruption because they do not measure it.

One key point in the present report is that evaluations of ACAs often rely on output indicators such as number of trainings conducted, number of prosecutions initiated, and so forth. These data are usually not systematically collected or readily available. Moreover, they only show whether the ACA has performed certain activities or not, revealing nothing about the results of these activities. Some agencies lack any system for tracking performance. The evaluations reviewed also have statements on performance at the outcome and impact level, but these are based on purely subjective and poorly substantiated analysis—that is, opinions. As a result, most assessments simply echo the negative stories already known to stakeholders. They provide no clear recommendations for future action, defeating the whole purpose of evaluation. Finally, most evaluations are conducted by external evaluators and consultants who are flown in for a few weeks to assess the performance of an agency. As explained below, this approach is highly problematic for a number of reasons.

This manual presents a methodology that gives ACAs and external stakeholders (including bodies formally tasked with the supervision of ACAs, such as civil society and international development partners) the tools to continuously monitor ACA performance at the outcome level, as well as to address issues of impact. By using such tools, ACAs can ascertain whether they are delivering on the expectations of their many stakeholders, building up public support, and reducing unjustified threats to their own existence. Following the minimum standards and the methodology suggested here should also produce higher-quality evaluations of ACAs, which can lead to a more informed debate on their effectiveness.

The methodology in this report has been developed and validated by experts, academics, and practitioners with wide-ranging experience in institutional design, development, and evaluation. The

primary users of the methodology are intended to include not only the ACAs and their national and international partners, but also external evaluators tasked with assessing ACA work. We do not seek to introduce a simple formula or evaluation/management technique that will ensure success whenever and wherever applied; it is precisely this form of oversimplification and one-size-fits-all approach that led previous interventions off track. Each agency is unique and requires a specific approach and set of indicators to monitor and evaluate its performance.

The proposed M&E framework may prove time consuming when first adopted, but it is cost-effective in the long term. A strong results-based M&E system aligns well with results-based management, which is beneficial for most public institutions. Donors and governments are becoming more and more cautious about granting further financial assistance to bodies with a poor performance record and/or low public support. ACAs that take up the challenge to improve their M&E systems and document positive progress and impact of their activities will likely find this redounds to their benefit. ACAs can better demonstrate their value to society if they can collect credible evidence of their performance.

If ACAs succeed in establishing clear indicators and perhaps even theories of change for their interventions, then evaluators will be able to do a better job. If the purpose of an evaluation is limited to accountability (answering the question of whether a particular ACA is using its funds well to produce outputs), then regular small assessments will suffice. However, if the purpose is learning as well as accountability, then evaluations that are able to answer questions about outcomes and impacts are needed. In other words, if we want to find out what works and what doesn't—and why—then we need more comprehensive evaluations over a longer time frame.

## 1.1 Organisation of the report

Following this introduction, chapter 2 explores how ACAs are currently evaluated through a mapping exercise of ACA evaluations and a critical overview of the existing methodologies and indicators that have been used to assess performance. This shows several shortcomings with respect to typical evaluation quality standards. This chapter also clarifies terminology and is relevant for all three audiences of the report: ACA staff, donors, and evaluators.

Chapter 3 discusses why it is important for ACAs to develop an M&E system and how conducting better evaluations can have a positive impact on public and donor support. ACA staff and donors are the intended audiences for this chapter.

Chapter 4 seeks to increase understanding of the evaluability of ACAs and suggest ways to improve it. The fact that ACAs have different institutional realities and face different challenges to their mandate influences their evaluability. To increase the evaluability of ACAs' work, we need theories of change (logic models) and development of indicators, M&E systems, data collection, and so on. The main audiences for this chapter are ACA staff and donors.

Chapter 5 outlines standards and basic methodological approaches for evaluations, seeking to address the shortcomings identified in chapter 2. Guidelines are provided for ACA staff, donors, and evaluators on how to design, plan, implement, and learn from evaluations. The overall approach recommended is one based on results and outcomes. This chapter is relevant for all audiences.

Chapter 6 considers key principles for the specific discipline of impact evaluation, including a methodology that can be used for future impact evaluations. This chapter is for readers interested in advanced evaluation methodologies.

Chapter 7 provides specific guidance on development of indicators for ACA evaluations. It introduces a function-oriented indicator catalogue (annex 1), with examples at the output, outcome, and impact levels for reference. ACA staff, donors, and evaluators would benefit from reading this chapter, particularly staff tasked with developing indicators.

Chapter 8 concludes the report by focusing on the important aspect of learning when conducting M&E, and provides guidelines for ACA and donor staff on how this aspect can be given higher priority. Evaluators would also benefit from reading this chapter.

## 2. Current practices in conducting evaluations

An extensive mapping exercise was undertaken by U4, the World Bank, and the United Nations Development Programme (UNDP). It covered most known ACAs, donor agencies, and consultants specialising in ACA work. The aim was to collect all possible assessments conducted on ACAs over the last decade. Surprisingly, this amounted to a rather short list. While the universe is limited, there were also stated reasons of confidentiality and a general fear that wide distribution of evaluations might discredit individual consultants, ACAs, or donors. This reduced the number of evaluations available for detailed consideration.

Thirteen evaluations were selected for close scrutiny. We found that their methodologies differed greatly, as did the quality and rigorousness of the exercises. To triangulate the findings, interviews were conducted with both evaluators and ACAs, probing their experience with assessing ACAs. Given the complexity of reforming public institutions, the sensitivity of the material, and the lack of prior relevant research, we opted to gather material from a range of sources.<sup>1</sup>

Based on this review, we can say that many of the accepted “truths” about ACAs have little solid research or systematic evaluations to substantiate them. Organisational capacity assessments are the most widely used form of evaluation, and in only a handful of the cases reviewed were the evaluations supported by even a “light” evaluation methodology. None of the reviewed evaluations used a methodology in line with internationally agreed evaluation standards (outlined below in chapter 5). Moreover, evaluations rarely tap data sources other than key informant interviews and document analysis, since M&E units internal to ACAs rarely produce such data.

This chapter clarifies the use of terms and explores how ACAs have been evaluated, the methodologies and indicators used to assess their work, the information left out, and the methodological problems involved. The focus of the evaluations varied depending on who initiated them. Some were meant to improve managerial decision making and strategy planning; others were the products of standard reporting mechanisms.

Most of the evaluations were initiated and conducted from the outside. They were carried out by external consultants on behalf of supporting donors, with the primary aim of monitoring the use of development assistance for accountability and control purposes. The resources dedicated to such evaluation, and the evaluators chosen, suggest that donors most often chose short, “light” evaluations, in which the evaluator was obliged to make subjective judgments about performance. The primary aim was rarely to produce evidence to show whether specific interventions are conducive to development. In the cases where this aim was stated, the time and resources were not available to build a methodology that would enable the evaluator to complete the task.

### 2.1 Understanding the terminology

Before presenting the analysis of the mapping exercise, this chapter clarifies the terminology used throughout the report. Since the evaluations reviewed were unclear in their use of M&E terminology, defining key concepts from the outset is important.

---

<sup>1</sup> The literature on ACAs is considered more or less in its entirety. Selected expert views expressed during an e-forum hosted by the newly created Asia-Pacific Integrity in Action Network (AP-INTACT) were also helpful. AP-INTACT has over 120 members, including country practitioners, international anti-corruption and public administration reform experts, and UNDP staff members. The discussion took place November 16–27, 2010.

First, although the words ‘project’ and ‘programme’ are used throughout this guide when referring to ACAs, this methodology does account for the fact that ACAs are organisations. There is no need to distinguish between an evaluation of a time-bound project and that of an organisation (or policy or programme). As noted by Gittinger (1982), a project is “the whole complex of activities for which money will be spent in expectations of returns.” We can thus consider ACAs as projects for the purposes of evaluation.

In the M&E field, confusingly, different terms are often differently interpreted by different audiences. Even the two main components, monitoring and evaluation, are not self-explanatory. Monitoring is an internal process within an ACA, a way to collect and analyse information on whether the ACA is on track to meet its objectives in a systematic and regular fashion. Evaluation, on the other hand, goes further and can be used to explain causal relationships and determine the value of a project, programme, or policy. Development practitioners often use the words “assessment” and “evaluation” interchangeably. In this manual, we adopt the view that a project assessment can include both appraisal and evaluation. Project appraisal assesses in advance whether it is worthwhile to proceed with a project. Project evaluation concerns itself with the performance of a project in a retrospective sense, that is, after it has been implemented (University of London 2010).

Many types of assessments and methodologies can be used to evaluate an ACA. Assessment tools for any kind of project—whether sponsored by a donor, government, nongovernmental organisation (NGO), or other entity - may be grouped into three main categories with respect to their timing in the project cycle:

- *Ex ante project appraisal:* Assesses whether funding should be provided to a particular project, through, for example, cost-benefit analysis or risk assessment.
- *Interim assessment:* Assesses performance while the intervention is being implemented so that programmatic adjustments can be made in response to monitoring data.
- *Ex post project evaluation:* Assesses the performance of the project after implementation, through, for example, qualitative or quantitative evaluation methodologies.

Such assessment tools are often mandatory for policies and projects in national governments of the developed countries, and donors also routinely use them (Boardman et al. 2006, 5; European Commission 2004, sections 4 and 5).

A classification of specific types of assessment tools is harder to develop, as no standardised list exists. However, five main types of assessments are covered in this manual:

- *Rapid assessment:* This focuses mainly on potential impact identification and screening as a form of “impact pre-assessment.” It is useful in situations where time is short and there is need for a prompt response, such as establishment of refugee camps (Glasson, Therivel, and Chadwick 2005, 284–86).
- *Process evaluation:* This is a variety of the regular programme evaluation (see next point) that focuses on the implementation and operations of a project or institution. It usually attempts to determine whether the project is aligned with its original design. The costs and time requirements for process evaluations are comparatively low.
- *Programme evaluation:* This is normally what is meant by the term “evaluation.” There are many different kinds of programme evaluations, but one commonality is that unlike assessments, evaluations do not make use of predefined standards, indicators, or criteria. Evaluations are used to answer questions related to design, implementation, and results. The evaluation design, method, and cost vary considerably depending on the type of question one

tries to answer. If cause-and-effect questions are being asked, the foundation for the evaluation is often a logic model or theory of change, which should be established at the beginning for each evaluation.

- *Impact evaluation*: This is a specific kind of evaluation with a sophisticated methodology that uses counterfactual analysis to attribute observed outcomes to the intervention. This is discussed in chapter 6.
- *Organisational capacity assessment*: So-called “performance assessments” in any field usually work by comparing the reality against predefined standards and criteria. As a result, these assessments are similar to audits, which assess the conformity of the intervention to procedures, norms, and criteria established in advance. They are not frameworks for evaluation. Organisational capacity “is the ability of an organisation to use its resources to perform” (Lusthaus et al. 2002, 15). The assessment of such capacities includes an examination of systems and management processes, as well as of human, financial, and infrastructure resources. As capacity development of partner country institutions is a core task for many development agencies, recent efforts have focused on trying to establish capacity assessments. It is important not to use these tools as substitutes for evaluations, given that capacity assessments have a different focus, that of exposing the gaps between desired capacities and existing capacities and thus identifying the needs of an institution for support.<sup>2</sup>

The above list is by no means exhaustive, nor are the types of assessments mutually exclusive. But they are the main types of assessments that can be used to assess ACA performance. Note also that global or regional anti-corruption review/self-assessment mechanisms are tools for compliance reviews and gap assessment, not evaluation frameworks.

Most of the reviewed assessments did not specify where they belonged in the above spectrum. Some claimed to be programme evaluations, but their approach and methods suggested that they were closer to rapid assessments or organisational capacity assessments. It was impossible to classify the assessments satisfactorily on this basis. Therefore the authors decided to denote all the assessments as “evaluations” in the discussion below, with the caveat that few actually qualified to belong to this category.

There is no right or wrong type of assessment. It all depends on what the purpose of the assessment exercise is. It is important to remember that the sophistication of the assessment methodology and the resources required are correlated. Our mapping of ACA evaluations showed that ACA evaluators were often given only a few weeks to conduct the whole evaluation. This time frame is normally suitable for a rapid assessment. Moreover, organisational capacity assessment methodologies were used for evaluations that were meant to capture changes at the outcome level, thus stretching the methodology beyond its intended use. Finally, any evaluation that is intended to measure outcomes of ACAs needs significant resources, planning, and support from the ACA staff and donors. This is even more the case when the evaluation is also intended to measure impacts.

---

<sup>2</sup> The classic sources are Lusthaus, Anderson, and Murphy (1995) and Lusthaus et al. (2002).

**Box 1. UNDP methodology for ACA capacity assessment**

The United Nations Development Programme (UNDP) has developed a “Practitioners’ Guide: Capacity Assessment of Anti-Corruption Agencies (ACAs)”, that is used to assess the existing capacities of a target ACA, keeping in mind the capacities that the ACA actually needs in order to discharge its mandate. The methodology is thus based on the classic organisational capacity assessment. The methodology has been designed to focus on *functions* performed by an agency, rather than the institutional arrangement or name of the agency. Thus, this tool can be used to assess the capacity of an anti-corruption commission with both prevention and law enforcement functions, an anti-corruption unit with a prevention function, an audit office with an investigation function or even a prosecutor’s office with an enforcement function. It is presented in a modular format to allow flexibility for ACAs to choose and focus on a set of key functions.

The results from the capacity assessment of ACAs provide the basis to develop and implement a comprehensive capacity development plan. This plan includes an integrated set of sequenced actions embedded in a programme or project to address the capacity development needs of a given ACA. The specific indicators and benchmarks established during the capacity assessment process can serve as a foundation for subsequent monitoring and evaluation of capacity development of ACAs. The assessment process is expected to mobilize and engage a range of actors involved in anti-corruption work in a given country. Moreover, the guide also provides a step-by-step guidance on how to carry out a capacity assessment exercise. While it can also be used in conjunction with other tools, this guide is not for evaluating outcomes or impacts.

*Source: (UNDP, 2011)*

There are also several types of monitoring, each with specific purposes:

- *Compliance monitoring*: Focuses on compliance with established standards and procedures.
- *Activity monitoring*: Focuses on quantity, quality, and timeliness of inputs and activities.
- *Financial monitoring*: Tracks cost of implementation according to budget and time frame.
- *Organisational monitoring*: Tracks sustainability, institutional development, and capacity building in the project and potentially directs attention to problem areas.
- *Context monitoring*: Focuses on the project context and environment, especially critical elements that can affect the implementation and progress of the project.
- *Beneficiary monitoring*: Focuses on the perceptions of beneficiaries of the intervention.
- *Results monitoring*: Focuses on achievement of planned results, especially outcomes and impacts.

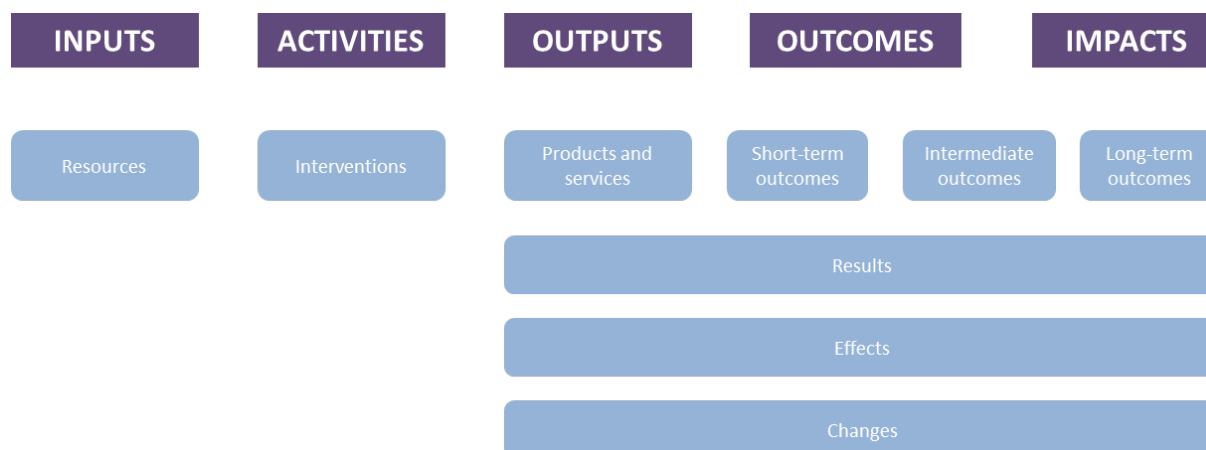
Most ACAs would benefit from implementing several of the above monitoring activities as part of their overall management approach. However, for M&E purposes, we emphasise the need for results-based monitoring (measuring outcomes and impacts) rather than only implementation monitoring (typically a mix of compliance, activity, financial, and organisational monitoring, measuring inputs, activities, and outputs).

In addition to the concepts of monitoring and evaluation, other terms in the broad field of M&E are at times subject to disagreement. The word “impact,” for example, has become very popular. It is often used synonymously and interchangeably with “results,” “outcomes,” and other performance markers.



In this paper we use the concepts of inputs, activities, outputs, outcomes, and impact as terms with distinct meanings. Following conventions of the Organisation for Economic Co-operation and Development (OECD), results can take place at three levels of a logical framework (logframe), namely outputs, outcomes, and impacts. In the figure below, other meanings or popular interpretations of these concepts are shown in light blue

**Figure 1. Clarifying concepts: From inputs to impact**



*Source: Adapted from slide presented at “Monitoring & Evaluation for Results,” training course organized by World Bank Institute Evaluation Group, Washington, DC, 2009.*

*Inputs* are the resources available to be used in a project, programme, or organisation, including money, staff, facilities, equipment, and technical expertise. *Activities* are what is done with the inputs, for example, holding seminars, producing manuals, etc. *Outputs* are the supply-side services or products produced as a result of a project’s activities. Examples might include training of 1,500 civil servants, implementation of a code of conduct, etc. *Outcomes* are the effects or results of the activities and outputs. Outcomes reflect uptake, adoption, or use of outputs by those who are supposed to benefit as a result of the project. Outcomes show changes, for example, higher levels of public trust. *Impacts* are the long-term consequences of a project, typically referring to goal attainment. Impacts are sometimes referred to as higher-level outcomes. An example could be “reduced corruption levels.” However, as shown below in the indicator catalogue (annex 1), it is important even at the impact level to specify which indicators are intended to capture the changes. No single corruption measurement tool or index can claim to measure “corruption levels” per se. Rather, they measure various aspects of or proxies for corruption levels, such as public perception, crime statistics, and so on.

“Impact” is a particularly hard word to define. Many existing discourses currently exist, some based on technical definitions, some on everyday use. A technical definition of impact evaluation is that it is a “counterfactual analysis of the impact of an intervention on final welfare outcomes” (IEG 2006, 1). In other words, impact evaluation may measure the extent to which the overall goal has been achieved. But it may also measure what change (impact) the intervention has produced compared to the likely scenario if the particular intervention had not occurred. Thus, when we describe below how we attempt to measure impact, this does not include all results, effects, or changes, but rather a specific long-term, ultimate outcome. While results can happen at the output, outcome, and impact levels, we use the term “results-based” to describe the M&E processes that emphasise the importance of measuring outcomes and impacts.

## 2.2 Results of the analysis

The framework for analysis was based on a set of quality standards developed by the OECD's Development Assistance Committee (DAC) in 2006.<sup>3</sup> The authors adapted these standards to produce the following 10 criteria for assessing the quality of ACA evaluations:

- Rationale, purpose, and objectives of an evaluation
- Evaluation scope
- Context
- Evaluation methodology
- Information sources
- Independence
- Evaluation ethics
- Quality assurance
- Relevance of the evaluation results
- Completeness

Given the scarcity of information provided in the evaluations, this mapping exercise could never make use of all these elements for any one evaluation. Among other reasons, the terms of reference for the evaluations unfortunately were formulated without these elements in mind. We provide guidelines for how to address this issue in chapter 5 below.

### 2.2.1 Rationale, purpose, context, and objectives

Almost all the evaluations set out to evaluate the effectiveness and performance of the institution in question. Few offered logframes with benchmarks against which progress could be assessed. In these cases, the suitability of the benchmark indicators was often also reassessed by the evaluators. Some also discussed the appropriateness and relevance of donor support to different institutional functions. Other assessments were meant to clarify the role of an ACA as an institution, or its role in the implementation of a national anti-corruption strategy. Only one evaluation prominently mentions the purpose of measuring impact.

Eleven out of the 13 evaluations reviewed were commissioned by donors supporting the respective ACAs. It is never plainly stated in the documents whether the donors paid for these evaluations or selected the consultants. However, some of these evaluations took place as part of a regular review (often annual or biannual) requested by donors by agreement with the ACA. In some cases, the evaluation was part of a memorandum of understanding between the government and several funding donors.

---

<sup>3</sup> The list of criteria presented here is based on the "DAC Evaluation Quality Standards" developed by the OECD/DAC Network on Development Evaluation for a three-year test phase application in 2006 (OECD/DAC 2006). The draft standards were revised based on experience and input from development partners and published by the DAC in 2010 (OECD/DAC 2010b).

### 2.2.2 Evaluation scope

The sample of 13 included evaluations of anti-corruption commissions, anti-corruption directorates under ministries, specialised corruption units under offices of attorneys general, and task forces to combat corruption.

In nearly half the cases, no information is given on the time taken to conduct the evaluation. Where information is given, evaluation normally took from 9 to 23 working days. This is remarkably little time given the multitude of tasks assigned to most institutions and the organisational and governance landscape that they must navigate.

None of the evaluations makes any reference to the costs involved. While funding partners might not be interested in revealing such numbers, this raises serious questions as to the openness about the methodologies used and particularly about their reproducibility. Ideally it should be established as a standard to publish the conditions under which an assessment has taken place and the methods used to draw conclusions.

Four of the 13 evaluations were conducted by a single consultant. In two cases the consultants were nationals. In the other two they were international consultants, both of whom were also involved in advisory tasks for the institutions under scrutiny. Four evaluations were conducted by teams of two. In three of these cases both consultants were international, while the fourth case involved two local consultants. Larger teams tended to be mixed, but the majority in each of the five additional cases were international.

### 2.2.3 Evaluation methodology

The majority of evaluations either did not have or did not specify a research methodology that went beyond key informant interviews and document analysis. This problem undermines the validity, reliability, and credibility of findings. Without a methodological foundation, so-called “light” evaluations are essentially just expert opinions. They provide little evidence that can be used confidently by policy makers or programme officers.

All but two of the examined evaluations noted that they were based on a desk review of relevant documents (literature, reports, project documents, international corruption indices) and interviews with stakeholders inside and outside of the institutions. Only two evaluations used secondary data to inform their analysis.

The majority of evaluations started by assessing the mandate and functions of the ACA. In addition, a number of evaluations also scrutinised the agency’s capacities to perform its functions, indicating what additional resources (usually staff) it needed to fulfil its mandate. In some cases, the organisational background (financial and human resources, audit and procurement systems, etc.) was addressed. The accessibility of the ACA and its services, the coordination of national anti-corruption strategies, and the cooperation with civil society were also evaluated in some cases. In general, the evaluations were oriented to outputs rather than to outcomes.

Nearly all evaluations alluded to the institutional or legal context of the ACA in question. However, this was generally only for the purpose of explaining the historical development of the ACA. While some evaluations make judgments on the independence of the ACA or its cooperation with other public institutions, few comment on factors critical for the ACA’s survival, such as political influence, the ACA’s own accountability mechanisms, and its management of expectations on the part of the public and decision makers. None of the evaluations contained an analysis of the political economy

surrounding the ACA. Only once was the question raised as to whether an institutional approach was the right one.<sup>4</sup>

Clear indicators that would be useful for measuring progress were found in only a handful of cases. This constitutes a major problem. If indicators do not exist or if they are not appropriate, then findings can be distorted. Typically, evaluators do not have time to construct indicators themselves, so it is up to ACAs and donors to make sure that appropriate indicators and result-based management systems exist.<sup>5</sup>

The existence of management work plans and logframes have pushed some agencies towards greater clarity on these issues. But even in these cases there is hardly ever an existing baseline, and often indicators leave room for interpretation. For example, the “number of investigations concluded in a year” can only tell part of the story, as investigations might take several years and are seldom started and completed in the same year. The “proportion of completed investigations leading to good prosecution or referral” leaves open the question of how to define a “good” prosecution. And lastly, the “increase in quality of investigations and prosecutions” will most certainly need more detailed follow-up indicators or criteria specifying quality.

For preventive functions, typical indicators are the existence of agreements with other public institutions, integrity committees, codes of conducts, and prevention policies in public institutions. The awareness-raising function is often measured only by numerical counts of campaigns conducted, radio programs aired, posters printed, and so forth. Little is known, as some evaluations point out, about the impact in terms of increasing citizens’ knowledge of corruption, much less changing behaviour.

For investigations and prosecutions, the usual indicators used are complaints received, ratio of complaints relevant to corruption to overall complaints, cases investigated, time taken to investigate, cases closed or handed to prosecution, cases prosecuted, and case acquittals. All these indicators are activity-based, with hardly any reference to an outcome or to the impact of an ACA on curbing corruption in a given area of risk. In some cases, more outcome-oriented indicators such as conviction rates and the recovery of assets are measured.

It is clear from the evaluations reviewed that more could have been done to point out the importance of indicators and well-constructed logframes. It appears that donor and ACA staff have developed indicators not for the purpose of evaluations, but rather for programme management purposes. This creates a fragile foundation for evaluation. The rest of the report presents suggestions on how logframes could be developed better in cooperation with the ACA and how they can be used by ACAs in their daily work as part of a results-based management instrument, as well as for evaluation.

#### 2.2.4 Information sources, independence, evaluations ethics, and quality assurance

In general, evaluators did not do a good job of providing information on the sources they used for the analysis, thus failing to adhere to fundamental evaluation standards. However, reading between the

---

<sup>4</sup> The Nepalese case study offered an interesting approach that merged contextual analysis with performance analysis. In taking into account the legal changes affecting the ACA’s mandate, the researchers examined its performance in several criminal cases and observed the changes from pre- to post-legislative reform.

<sup>5</sup> For instance, a 2007 review by the UK Department for International Development (DFID) of support to the Sierra Leone ACC found so little progress that termination of the support was suggested. A 2008 DFID review of support to the Malawi Anti-Corruption Bureau found good progress, which was, however, poorly communicated. A 2010 annual review of the same institution found that the targets had been surpassed and were in need of readjustment.

lines, it is clear that the primary information sources in the majority of cases were document reviews and interviews with key informants.

Unfortunately, only a handful of evaluations provide lists of documents reviewed and persons interviewed. This undermines data credibility. The evaluations that provide such information refer to a range of different stakeholders who were interviewed, although the overall number and balance seemed limited by time constraints.

This leaves significant question marks regarding the independence and objectivity of the evaluations under review. Do they represent an honest account of the state of affairs? Do evaluators tend to be positive when they feel the client wants a positive result and negative when that is seen as the preferred option (for example, when the intention is to cut funds)? In other words, are they objective or biased?

It cannot be directly deduced from the evaluations which of them were paid for and potentially influenced by donors. However, in-depth interviews for this study seem to suggest that donor influence towards a preferred outcome is common. Experienced evaluators and former commissioners suggest that donors get the results they want, whether positive or negative. While this is essentially a political issue, there are several ways in which a stronger evaluation methodology could limit the scope for subjectivity and discretion and thus reduce the likelihood of biased conclusions, for example through established standards and quality assurance processes. These are described in chapter 5.

On quality assurance, one can assume that many of the donor-supported evaluations were shared with the agencies in an ad-hoc manner. Allowing the institution under evaluation to comment on the findings is not only a matter of transparency and good ethical standards; it also reduces the margin of error, increases data quality, and validates the conclusions. However, the extent of this practice is not clear from the review of the evaluations. Donors and ACAs should work together to promote better and more transparent quality-assurance processes. Possible approaches are suggested in chapter 5.

Nearly all the evaluations reviewed failed to note whether an internal ACA M&E unit was involved in the process. Whether such units were nonexistent or too weak, it is striking that the recommendations made no mention of the need to strengthen them and include them in future evaluations. This is remarkable, since such units should be central in producing data on key indicators and tracking progress between evaluations. In most cases, they could also serve as the champion for performance and outcome/impact assessment, as well as communication.

### 2.2.5 Relevance of the evaluation results

This criterion is perhaps the most important of the 10, since it focuses on the basic value of the evaluation for the stakeholders and beneficiaries. For evaluations to have real relevance for ACAs, the learning elements must be emphasised (see chapter 8 for a broader discussion).

On a positive note, we find that some evaluations discuss to what degree the recommendations of earlier evaluations have been taken up. However, interview material seems to suggest that externally induced evaluations are given to senior management and do not penetrate further into the organisations. This suggests that evaluations are used more as external accountability tools than as learning tools. If one wants to maximise the learning potential of an evaluation, it should be widely distributed. It was beyond the scope of this exercise, however, to find out how the ACA staff members regard the relevance of the evaluation and to what degree they benefited from its recommendations.

## 2.3 Benefits of better evaluation

On the basis of this brief review of how evaluations of ACAs have been conducted, we are convinced that current methodologies employed by the donor community can be improved. Better evaluations of

the outcomes and impact of an ACA's work would provide donors with a firmer evidence-based foundation on which to decide whether to strengthen or cut off support for the ACA. Such advances would also be useful for evaluators, who currently have no strong methodologies on which to base their work. ACAs would be the primary beneficiaries, as their own working systems would be strengthened, they would be better able to learn from their efforts, and they would ultimately do a better job fighting corruption.

### 3. Why M&E matters for ACAs

Any public agency should be accountable for its use of public resources. This requires that a performance evaluation system be in place. Because of their unique mandate, ACAs should also be able to demonstrate their effectiveness and impact to the citizens of the countries where they operate.

Most bold statements or “stories” about the effectiveness of ACAs are not corroborated by strong evidence. In order to be prepared to make their case heard in what is increasingly a competitive environment for donor and national funding, and one where many question the necessity of governance work, ACAs must make a commitment to take M&E seriously. They need to go beyond current practice and document their achievements with solid evidence.

This requires, first, a clear definition of what is meant by M&E. *Monitoring* refers to “a continuous function that uses the systematic collection of data on specified indicators to provide management and the main stakeholders of an ongoing development intervention with indications of the extent of progress and achievement of objectives and progress in the use of allocated funds” (OECD/DAC 2002, 27). *Evaluation* means “the systematic and objective assessment of an ongoing or completed project, program, or policy, including 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 lessons learned into the decision-making process of both recipient and donors” (OECD/DAC 2002, 21). Thus, monitoring and evaluation are interlinked but independent exercises, with different processes and purposes.

Given the investments made in ACAs, one would think that M&E would be an integral component of operations. Our findings indicate otherwise. For most ACAs, therefore, it is necessary to take one step back and get the basics in order before attempting any advanced evaluation efforts. It is of vital importance that an M&E framework be established at the outset, specifying targets and time frames in advance. ACAs should take the time to develop good indicators, including at the outcome and impact levels, and learn how to communicate their results (De Sousa 2009a, 2009b).

Targets should be agreed among the ACA, the government, donors, communities, and civil society organisations to ensure buy-in and mutual accountability. Sudden changes in targets, that is, moving the goal posts, should not be allowed.

Organisations tend to resist the need for monitoring and evaluation of their performance. It is seen as an expensive and tedious exercise that gets in the way of doing “real work.” Pressures for M&E cite two main purposes: accountability and learning. The pressure for accountability is usually what leads to M&E work on donor-funded projects. Donors need to know how their money is being spent and how well it is spent. However, M&E is also a tool for learning. As mentioned, results-based management systems are based on the same principles of collecting information about the performance of activities. This in turn improves the “real work” of the ACAs.

#### Box 2. Why is it important to measure results?

- If you do not measure results, you cannot tell success from failure.
- If you cannot see success, you cannot reward it.
- If you cannot reward success, you are probably rewarding failure.
- If you cannot see success, you cannot learn from it.
- If you cannot recognise failure, you cannot correct it.
- If you can demonstrate results, you can win public support.

*Source:* Osborne and Gaebler 1992.

M&E enables ACAs to highlight the benefits, risks, and limitations of their activities. It is not merely something that has to be done in order to showcase a project's accomplishments for the benefit of donors and taxpayers. M&E is fundamentally a learning process, one that can represent a shift in the way ACAs operate and reflect on the impact their work is having in the surrounding context.

M&E reports often reveal mistakes in the way a certain investigation or inquiry was conducted. Evaluating how an awareness-raising campaign was promoted or how a complaints system was implemented provides a basis for questioning underlying assumptions and "normal ways of doing things," and offers paths for innovation and improvement. M&E exercises which emphasize learning are thus also important capacity-building activities. They can establish useful routines, activities, and motivation for M&E within an organisation, so that it can become a learning organisation.

Establishing internal M&E units is beneficial at many levels, as it promotes both accountability and learning. Limiting evaluation to external evaluations where international experts are "parachuted" in for a few days, in contrast, completely bypasses the learning element in M&E.

### **Box 3. Why is M&E important to ACAs' work and existence?**

*Transparency:* Visibility and outreach are important in attracting public support. Public ignorance about the existence and functioning of an ACA lays the conditions for its marginalisation or gradual death. M&E generates written reports and concrete performance figures that contribute to transparency and visibility.

*Accountability:* ACAs are publicly funded bodies. Therefore they need to report on their activities, capacity problems, and results to those who fund their activities, that is, taxpayers and donors. M&E provides reliable information on performance and helps to track progress more easily and systematically.

*Institutional memory:* M&E enables governments and donors to ascertain whether an ACA is able to fulfil its mandate by tracking its performance and growth in capacity over a period of time. It also allows the ACA to develop an in-house memory about the different phases of institutionalisation.

*Learning:* M&E provides solid evidence for questioning and testing assumptions and integrating important lessons and experiences into policy and practice. It offers a basis and a process for self-reflection. Finally, it also provides ACAs with a more robust basis for raising funds and influencing policy.

*Improving policy:* Strong M&E frameworks give the heads of ACAs and governments indications of whether a policy option is working as intended, by detecting operating risks and problems. Where do the problems originate? How is the agency's performance affected? What capacities/resources are available to reduce those risks and problems and can those be strengthened?

*Better performance:* All of the above should lead to better performance of ACAs in fighting corruption.



## 4. Challenges in ACA evaluability

This chapter serves two purposes. First, section 4.1 aims to clarify the concept of ‘evaluability’ of ACAs. Second, on the basis of this understanding, section 4.2 formulates strategies for increasing the evaluability of ACAs.

M&E in the field of anti-corruption is underdeveloped. Possible reasons for this include lack of attention to the importance of evidence and measurement as well as lack of technical competence in the field to conduct “professional” evaluations. Or perhaps it is simply a symptom of the widespread belief that the area of anti-corruption presents “special” challenges which make rigorous evaluations unfeasible and futile.

The sections below present some of the common challenges in the field, discuss what institutional characteristics evaluators must take into account, and briefly outline what methods can be used to increase the evaluability of ACAs. We conclude that ACAs are not easy to evaluate, but that they are “evaluable” if certain elements are developed internally within the ACAs. Credible findings can be produced if a strong methodological approach, results chain, and indicators are developed and used, and if care is taken in collecting and analysing data.

The fact that ACAs are multifunctional organisations which pursue different targets in an area where data are not always easy to obtain should not lead to the conclusion that they can only be assessed through so-called organisational capacity assessments. Similarly challenging multifunctional organisations, such as the Social Funds for Development,<sup>6</sup> exist in other sectors where focused work on developing internal M&E capacity has increased their evaluability.

This chapter is relevant for all three target groups: ACA staff, donors, and evaluators. ACA staff members need to know which elements of their internal systems they should strengthen before strong evaluations of their work can be done. Donor staff benefit from knowing the criteria that determine when it makes sense to conduct an evaluation of an ACA, and from understanding how better evaluations can be done. Evaluators need to be informed about the specific institutional characteristics of ACAs they will have to take into account, the challenges they will face in this field, and the methods they can use to overcome the challenges.

### 4.1 Understanding the evaluability of ACAs

Most of the evaluations of ACAs reviewed as part of the mapping exercise for this study should not have been undertaken, as the findings were based on little evidence and were deemed to be subjective. This is not the fault of the evaluator. Rather, it reflects a lack of recognition that certain crucial elements need to be in place before the evaluator can do a proper job. In other words, a number of preevaluation activities must be undertaken before an evaluator can be sent in: clear goals/objectives should be in place, performance indicators should be set up, an M&E system should be operational, and so on.

The concept of “evaluability” refers to the extent to which an activity or a programme can be evaluated in a reliable and credible fashion. The assessment of evaluability involves an early review of the proposed activity in order to determine whether its objectives are adequately defined and its results can be verified (OECD/DAC 2000, 2009). It also reveals barriers that might impede a useful evaluation. Evaluability assessment “requires a review of the coherence and logic of a programme,

---

<sup>6</sup> The impact evaluation of the Social Fund for Development in Yemen is one example (Recovery & Development Consortium 2010).

clarification of data availability, an assessment of the extent to which managers or stakeholders are likely to use evaluation findings given their interests and the timing of any evaluation vis-à-vis future programme or policy decisions” (European Commission 2009).

If a project is not ready for evaluation, recommendations can be given on how to make it ready for such a process. Frequently the shortcomings of the programmes are connected to three areas: programme design, availability of relevant information, and conduciveness of the context.<sup>7</sup> Various handbooks and checklists offer advice on how to assess evaluability.<sup>8</sup>

#### 4.1.1 Common challenges for measurements in the field

As observed by Fredrik Galtung, “until the mid-1990s, most of the empirical findings on corruption in the academic literature were of an incidental or anecdotal nature” (2005, 101). In a review of the state of research on corruption, Bardhan wrote candidly in 1997 that “our approach in this paper is primarily analytical and speculative, given the inherent difficulties of collecting (and hence nonexistence) of good empirical data on the subject of corruption” (1997, 1320).

Social scientists are faced with two important questions when they want to study corruption:

How does one study something that is defined in part by the fact that individuals go to great lengths to hide it? How does one accurately measure the extent of corruption when attempts at measurement may cause the actors involved to either reduce their illicit behaviours during the periods of measurement, or find new ways to obscure their behaviours? (Banerjee, Hanna, and Mullainathan 2009, 1)

Most anti-corruption practitioners adhere to the view that “the science of measuring corruption is more an art form than a precisely defined empirical process” (UNDP 2008, 3). The challenge for measuring the fight against corruption is that off-the-shelf toolsets, indicators, and reliable statistics are hard to produce because of the very context-specific, clandestine, underground nature of corruption (Langseth 2006, 16).

This litany of complaints is not unique to the field of anti-corruption. Many fields consider themselves to be “extra challenging.” Despite such difficulties, a range of methods are available today which can produce solid information and data to evaluators in the field of anti-corruption. These are described in section 5.4.

Nevertheless, the question remains as to whether one needs to measure levels of corruption to ascertain the performance of an ACA. The starting point is to clarify whether corruption levels are indeed the key variable for performance. Many ACAs may have a bigger role in coordinating corruption policy activities than in directly fighting corruption. If ACAs are tasked with directly fighting corruption, they often focus on a specific type of corruption, such as grand corruption, political corruption, or public sector corruption.

If one is performing a regular programme evaluation, it is worth emphasising that many of the OECD evaluation criteria do not necessarily correspond to corruption levels. The “effectiveness” and “efficiency” of an ACA would be evaluated by assessing internal ACA processes, outputs, and resources used, comparing them against those of similar organisations. Such measures are not dependent on measurement of corruption levels.

---

<sup>7</sup> UNIFEM (2009) has a useful checklist on pages 3–4.

<sup>8</sup> See, for example, Wholey, Hatry, and Newcomer (2004, 33–41).

The “impact” criterion is more problematic. In order to understand an ACA’s impact, one typically has to measure some kind of change in levels of corruption (depending on the mandate, tasks, and focus areas of the ACA). Any public agency has overall objectives, and in order to understand performance it is necessary to measure changes in relation to these objectives. The typical and problematic response is to fall back on generic macro-level indicators, such as Transparency International’s Corruption Perceptions Index (CPI), because these are seen as the only impact-level data available. However, such very broad indicators rarely capture well the essence of the desired impact of an ACA.

An example from another sector can illustrate this point. In an evaluation of a health programme delivering vaccines to children, it is not solely the longevity rate (an overall impact indicator for health) that is measured. Many people would find other aspects of health equally or more important, such as quality of life, absence of severe diseases, etc. Moreover, for such a programme, the desired impact would relate specifically to children as a target group. Health is a concept that is defined differently depending on the intervention one is evaluating. The same is true of anti-corruption. Nationwide perception levels or control-of-corruption indexes may be informational at a general level. But they only tell part of a wider story. They should not be taken as substitutes for project-level or organisation-level measurements.

Is it then impossible to measure impact of anti-corruption activities? The short answer is no. However, when designing and implementing programmes, it is necessary to identify the specific output, outcome, and impact variables one needs to measure. Evaluators should use common sense when dealing with impact questions. Reducing corruption levels in a country is a major undertaking which requires significant resources and inputs from a myriad of national integrity actors. One can compare it, in the level of generality, to improving health or reducing poverty in a country. Would an evaluation ever hold a specific health- or poverty- reduction programme or organisation accountable for the general levels of health and poverty in a country? Sometimes the answer might be yes, but only if these programmes had significant resources and if measurements were taken over a very long time frame. For smaller programmes over the medium term, it is more fruitful to unpack overall concepts such as corruption into more specific indicators and base the measurements on these indicators.

#### 4.1.2 Different institutional characteristics of ACAs

The world of ACAs is more colourful and diverse than one might expect. We therefore warn against the temptation to develop standardised evaluation methodologies or “toolkits” intended to be applied uniformly to all ACAs. In general, the design of an ACA should not affect its evaluability, but evaluators should be aware that the intervention logic or theories of change can differ greatly from one ACA to another. Thus, evaluations must understand and take into account the ACA’s specific context and challenges. Some countries have decided to set up new institutions to fight corruption; others have simply renamed or reviewed the competences of existing ones. Some have preferred to set up ACAs with a preventive, educational, and informative role. Others have gone further by establishing multipurpose specialised ACAs with a mandate to prevent, investigate, and in some cases prosecute corruption.

The literature identifies a series of prerequisites for a particular agency to be classified as an ACA (De Speville 2000; Doig 1995, 2000; Pope and Vogl 2000; Pope 1999; Quah 2000). In practice, however, very few ACAs have fulfilled all these requirements. There is no standardised model of what an ACA should look like. Some have been newly created as statutory bodies. Many are constituted as special units in prosecution and ombudsman offices, national audit agencies, or police forces. Many are also set up with purely preventive capacities or coordinative roles, a limitation which can lead to speculation that the institution was established merely to satisfy donor demands.

In a recent report entitled *Specialized Anti-Corruption Institutions: Review of Models* (OECD 2007), the OECD attempted to classify existing agencies into three types:

- *Specialised institutions with multiple competences:* This refers to anti-corruption agencies *sensu stricto* with preventive and enforcement powers and a wide spectrum of activities that go beyond traditional criminal investigation, such as policy analysis, counselling and technical assistance, information dissemination, ethics monitoring, training, and scientific research (risk assessments, surveying, etc.).<sup>9</sup>
- *Specialised departments within police forces or prosecution bodies:* In some cases, the same body combines investigation and prosecution competences, such as the Romanian National Anti-Corruption Directorate (DNA). It is difficult to establish to what extent this type differs from the other two, as some of these bodies also have preventive functions.<sup>10</sup>
- *Institutions with exclusively preventive competences:* These institutions may conduct scientific research on corruption, develop and advise on corruption control policies for decision-making bodies, monitor and recommend amendments to regulations in high-risk sectors in the public and private spheres (e.g., urban planning, public works, state procurement, casinos, customs authorities, etc.), monitor the rules on conflicts of interest and declarations of assets, draft and implement codes of conduct, assist public servants on corruption matters, facilitate international cooperation in this field, and act as intermediaries between civil society and state bodies with competences in the area.<sup>11</sup>

One obvious but sometimes overlooked issue in measuring ACA performance is that the powers and capacities of the ACA should be aligned with the expectations placed on it. For example, an ACA with purely preventive competences should not be measured using law enforcement indicators. This point will be elaborated further in chapters 5 and 6.

Evaluators should take note of a number of institutional characteristics and external factors which typically have great influence on the performance of an ACA. Capacity issues that can result in institutional failure at multiple levels include:

- *Contextual factors:* The degree of success of ACAs must be seen in context. Organisational cultures, levels of national development, and political stability set the background scene for these bodies. It can hardly be expected that an ACA will function adequately in a country with severe governance problems (De Speville 2000; Doig, Watt, and Williams 2007).
- *Legal framework and policy factors:* ACAs encounter various constraints to their mandate as a result of policy choices during the legislative process, which may determine their legal status, institutional location, special powers, sharing of competences and information, financial autonomy, reporting procedures, etc. Under their statutes, all ACAs are in some sense “independent.” In practice, however, the degree of operational autonomy varies considerably

---

<sup>9</sup> Agencies of this type include the Hong Kong Independent Commission Against Corruption, the Singapore Corrupt Practices Investigation Bureau, the New South Wales (Australia) Independent Commission Against Corruption, the Botswana Directorate on Corruption and Economic Crime, the Special Investigation Service of the Republic of Lithuania (STT), and the Latvian Corruption Prevention and Combating Bureau (KNAB).

<sup>10</sup> Other examples include the Norwegian National Authority for Investigation and Prosecution of Economic and Environmental Crime, the Central Office for the Repression of Corruption in Belgium, the Spanish Fiscalía Anticorrupción, the Central Investigation Directorate on Corruption and Economic and Financial Crimes of the Portuguese Judiciary Police, and the Central Department for Penal Action and Investigation of the Portuguese Attorney General’s Office.

<sup>11</sup> Examples include France’s Service Central de Prévention de la Corruption (SCPC), the State Commission for Prevention of Corruption in Macedonia, the Albanian Anti-Corruption Monitoring Group, the Permanent Commission against Corruption in Malta, and the United States Office of Government Ethics.

from one agency to another. In many cases, these bodies are operationally and financially independent in name only.

- *Organisational factors:* Low levels of performance also derive from inadequate recruitment and accountability procedures and inadequate or nonexistent management arrangements necessary to the organisation's capacity to operate and deliver on its mandate. The technical capacity of an ACA can also be hampered by ineffective collaboration with other competent authorities. Difficulties in obtaining evidence about corruption practices or information about risk areas from other state bodies or agencies reduce the effectiveness of ACAs. ACAs also often lack the authority to ensure that their recommendations are enforced by public institutions. In most cases, the legal framework for inter-institutional collaboration is not carefully addressed at the outset. In addition, ACAs often find themselves stretched to deliver on overly broad mandates. Those that are multipurpose agencies end up in the dilemma of being expected to clamp down on corruption (a fear-based function) while at the same time serving as an advisory body on prevention (a trust-based function).
- *Financial factors:* Lack of financial resources is a constant threat to any organisation. While large budgets do not necessarily generate greater levels of productivity, it is important to note that some ACAs work under strained conditions that may seriously compromise the pursuit of their objectives.
- *Leadership and expertise factors:* The individual skills, experience, and knowledge of ACA staff are fundamental to their success. Capacity issues concern both the technical capacities of ACAs and their overall functional capacities (such as leadership, human resource management, planning, organisational learning). One of the advantages of ACAs in comparison to traditional law enforcers is their capacity to generate a knowledge-based approach to the fight against corruption through risk assessments and a series of other studies. In principle, these bodies ought to be provided with a team of experts (while also drawing on the knowledge and experience of other monitoring and regulatory units and sharing their own in exchange). In practice, however, very few have developed such in-house research and knowledge-production capacity.

These different capacity issues pose different challenges to M&E and to the issue of evaluability. In short, ACAs have certain commonalities which an evaluator can rely on. But it is also crucial to recognise the subtle differences in mandates and functions and the factors that typically account for institutional failure. This can make it possible to increase the evaluability of ACAs and promote evidence-based evaluations.

## 4.2 Increasing the evaluability of ACAs

An ACA's programme and institutional design should have features built in to ensure its evaluability. It is often too late to remedy key issues if one waits until the commissioning of an evaluation. Several factors can increase the evaluability of ACAs. Many of these depend on involvement from ACA and donor staff and on the availability of resources to build capacity within ACAs in the area of M&E. Where capacities are below standard, ACA staff should work on a number of activities (possibly with the support of donors and experienced trainers/facilitators) before the evaluation commences.

A European Commission (2009) sourcebook on evaluability assessments recommends five steps to increase evaluability:

- Make explicit the intervention logic and results chain, i.e., clarify the high-level objectives (goals), intended mechanisms and agents, expected outcomes, and assumptions underlying the intervention.

- Judge the extent to which the design, strategy, resources, and implementation mechanisms are appropriate given the intervention logic.
- Judge the extent to which the programme is likely to lead to the goals of the overarching policy or priority for which it was set up.
- Assess the availability of necessary information, including primary and secondary data sources and the likely ease and cost of access, including the willingness of gatekeepers to cooperate.
- Consider how far an evaluation is likely to lead to real improvements in programme performance and success and whether this is commensurate with the likely costs and efforts of undertaking an evaluation.

The first three steps focus on establishing clear theories of change and performance indicators for ACAs. Guidance on how to do this is given in section 4.2.1 below. The last two steps depend on the ACA having a functional M&E system in place, and on the extent to which findings will promote learning. Section 4.2.2 provides guidelines for this.

It should be noted that these exercises might be part of the evaluator's job: in other words, the evaluator could be asked to establish a theory of change and indicators as a first part of the assignment. In these circumstances, sections 4.2.1 and 4.2.2 would belong under chapters 5 and 6, which deal with how to better measure outcomes and impact. However, it is usually preferable to have strong involvement by ACA and donor staff in the construction of theories of change, indicators, and the M&E system, as they have more intimate knowledge of the intervention logic. This can also serve as a capacity-building exercise. The evaluator should serve as a facilitator and trainer rather than taking primary responsibility for these steps.

#### 4.2.1 Establishing clear theories of change and performance indicators

This section focuses on the importance of establishing goals and lower-level objectives for the ACA, and on identifying ways to measure whether the organisation is on the right path to achieve its goals. This is one part of a broader results-based management approach, but this section focuses particularly on the M&E aspects. If theories of change and good indicators are established, this substantially increases the evaluability of the ACA and the chances for better-quality evaluations. This section is therefore also relevant for chapter 5 below, on how evaluations can be performed better.

First, we present some common challenges for ACAs in establishing logic models and theories of change (ToCs) and performance indicators. Then we provide guidance on how to establish good ToCs and performance indicators for ACAs.

##### *Common challenges for ACAs in establishing goals, theories of change, and indicators*

Very few ACAs define the goals and expected outcomes and outputs of their work in their strategic plans. This makes it difficult to assess how successful they have been. As Meagher puts it, "The extent to which the objectives of a new agency reflect a desire for systemic change, as opposed to a drive to score political points, is rarely clear—and is rarely intended to be clear. This poses a dilemma for the observer in defining and measuring success. What yardstick to use?" (2002, 7).

It is advisable for ACAs to disaggregate their goals as much as possible. Rather than having a goal of reducing corruption nationwide, it would be preferable to break down this goal into more specific and observable units. One way to do this is by specifying the *arenas of corruption*. For example, is the focus on the public sector or the private sector? If it is on the public sector, are efforts aimed at the central or decentralised levels of government? If it is the central government, are particular ministries being targeted? It is also advisable to distinguish between the *types of corruption* targeted:

bureaucratic versus political, petty versus grand, etc. Theories of change and indicators flow from the goals. If the goals are too broad or only represent political signalling rather than realistic targets, then evaluators are given a difficult, if not impossible, task.

Usually, ACAs have broad goals that apply nationwide, making it difficult to establish discrete beneficiary groups. In addition, the typical goal of contributing to a reduction in corruption is inherently difficult to measure. Even if a reduction in corruption can be established, it is another matter to attribute it in a causal chain to an ACA's work, especially since ACAs usually are not the only actors in a national integrity system. As Meagher (2002, 10) explains,

An ACA's record of success—how well it carries out its tasks—does not, by itself, measure net value-added. If the ACA didn't exist, some other agency would likely be handling its preventive, deterrent, and outreach functions. Do ACAs do this better—so that they outperform other arrangements in producing these outputs and outcomes? If so, how do they do it, and how do we know?

Thus there is an attribution problem as well as a problem in demarcating the target group. In some cases, moreover, goals are not formally stated, or the formal goals may be significantly different than the actual intended impacts.

ACAs often use outputs as proxies of outcomes, but they are poor substitutes. An output measure, such as the number of cases investigated, can in no way be considered indicative of an outcome. It only measures levels of activity, not accomplishment or change. When using quantitative output indicators, it is also important “to apply a qualitative sense of whether the agency and activities are well targeted, and whether the outcomes are as beneficial as they could be” (USAID 2006, 24).

As with other public institutions, the survival of ACA depends on factors other than performance standards. ACAs “are created and terminated by political decision and not always an informed one” (De Sousa, Hindness, and Larmour 2009, 14). In addition, they are more often than not kept under tight control by the government and even used as political instruments in numerous ways (such as to target the opposition or to please the public and the international community). Such hidden agendas further complicate assessment. One important way that an evaluation can compensate for the politicized aspect of ACAs, although not completely, is to link performance expectations to availability of resources. Agencies with insufficient and/or low-skilled staff, poor resources, and no legal competences to pursue criminal offences cannot be expected to produce significant outcomes.

All these obstacles make evaluations difficult. An evaluation depends on having clear strategic objectives and identified goals, purposes, and outputs, which lay the foundation for the evaluation. If goals, objectives, and indicators are too wide, general, unclear, or simply nonexistent, it is difficult to undertake an evaluation. This is also the case if the stated goal is not the same as the actual goal defined by political realities. ACA managers need first to identify the true vision and mission of the organisation, with associated goals and indicators, before evaluations can be undertaken productively.

#### *Mapping out the logic model/theory of change for the ACA*

A logical, coherent, consistent, and clearly articulated theory of change underlying the ACA's interventions should be in place before any evaluation is undertaken. This can be shown in a logic model or results chain. This is one of the fundamental building blocks of our proposed impact evaluation methodology (chapter 6). Making use of the terminology defined above, we can say that organisations implement programmes to achieve a desired *impact or impacts*. Programmes use *inputs* and *activities* that result in *outputs*. *Outputs* should lead to desired *outcomes*. In the longer term, *outcomes* will be reflected in *impacts*.

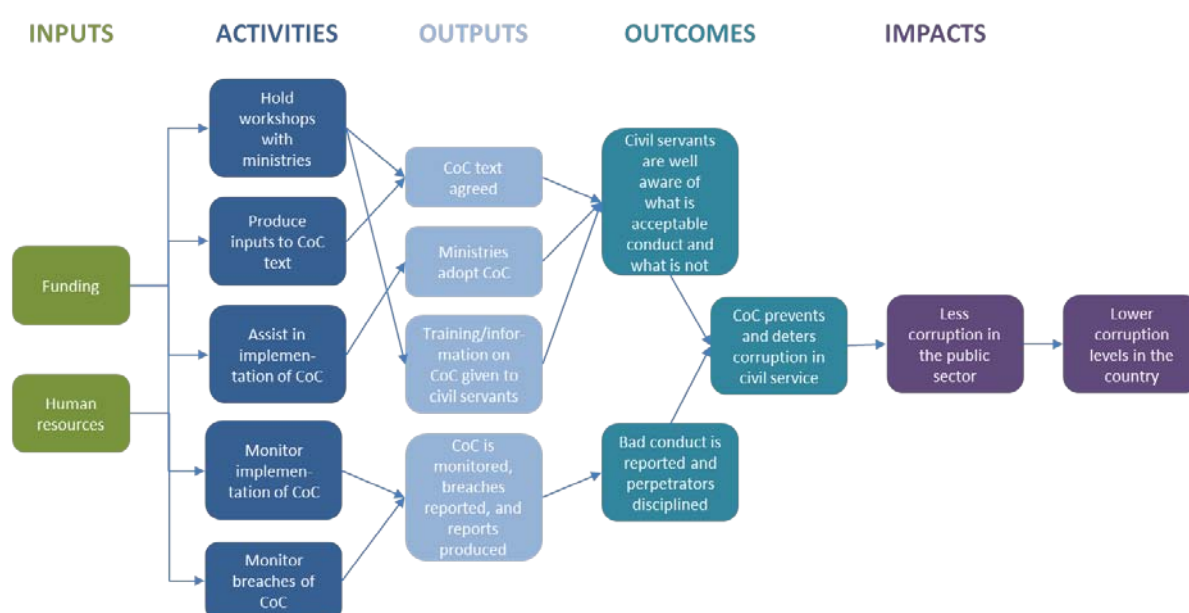
As can be seen from figure 2, there should be logical links between all inputs, activities, outputs, outcomes, and impacts. If outputs are produced which do not contribute to any outcome, they are either superfluous or the model is not constructed properly. This same logic lies behind well-constructed logframes: outputs lead to outcomes (purposes), and outcomes lead to impacts (goals).

In cases where a logic model is not in place, the evaluator may construct a proxy logic model and get approval to evaluate based on that. However, this is not recommended.

It might be necessary to construct several logic models for a single ACA, given that ACAs are often multi-function, multi-objective organisations. Generally, if the activities undertaken by the ACA have very distinct outcomes and/or impacts, then one should consider constructing separate logic models, as the ToCs underpinning the various interventions would be different.

To illustrate, a hypothetical example is provided of how a logic model might look if the ACA were responsible for implementation of a code of conduct (CoC) in the public sector.

**Figure 2. Theory of change for implementation of a code of conduct**



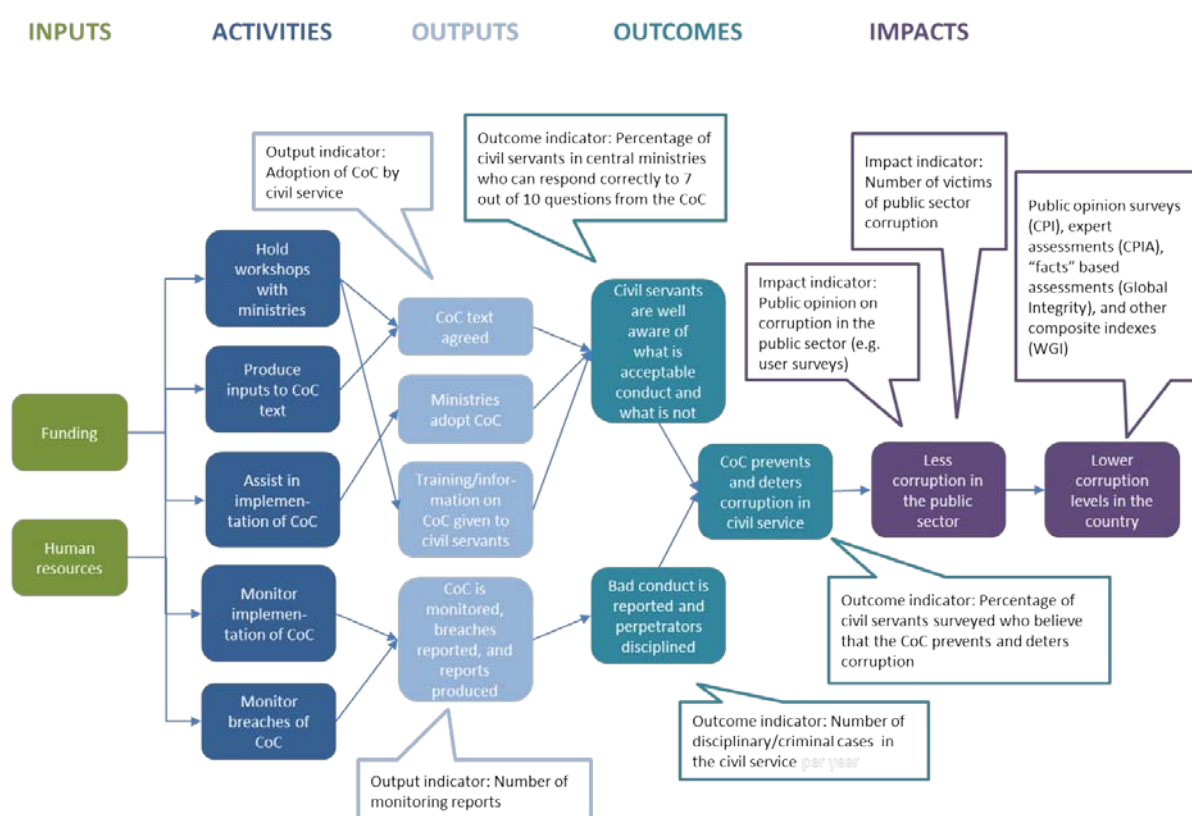
This logic model shows how working with a code of conduct can lead to the desired impact of less corruption. It clarifies the theory of change by specifying that the code of conduct is expected to have both a preventive and an enforcement effect. Several of the assumptions behind this ToC can be questioned. For example, it is not certain that awareness of standards for proper conduct and of the sanctions for breaching them would in fact deter people from engaging in corruption. This is highly context-specific. However, outlining the assumption in an explicit ToC allows it to be examined. There is no standard model for a valid ToC. But in principle, it should reflect the working mechanisms of the intervention as well as possible, be developed in a participatory manner, and be agreed between stakeholders.

For programme planning purposes, construction of a ToC is a hypothetical exercise that aims to anticipate the best possible way in which the intervention might have the desired effects. When a ToC is being used in an evaluation context, it should approximate reality as closely as possible, making possible an explicit comparison with the previous hypothetical ToC constructed in the design phase. Although a ToC is often not formulated explicitly, there is always some rationale behind any purposeful intervention.



Once the logic model is constructed and a ToC has been mapped, it is possible to decide which key parameters to evaluate. Indicators can be constructed for each key parameter, and the logic model makes it easy to identify whether indicators are at the output, outcome, or impact level. Figure 3 shows examples of indicators corresponding to the different levels of the logic model. Note that up to and including the outcome level, the indicators are specific to the intervention (in this case, implementing a code of conduct), but at the impact level they become more general and less specific to the intervention. Thus, when evaluating an ACA one cannot only measure impact indicators, as these do not show whether the impact was caused by a successful ACA intervention or by an external event or other organisation. In figure 3, one of the impact indicators is whether users of public services have a more positive opinion regarding levels of corruption in the public sector. This indicator could have become positive because the government just installed a new whistleblower hotline, in which case it would have nothing to do with the code of conduct. Only if we also measure the outcome indicators can we begin to understand whether the code of conduct has made a contribution to reducing overall levels of corruption.

**Figure 3. Theory of change with indicators**



The need for indicators that are specific to the intervention also requires that indicators be constructed, and data gathered, at the project or organisational level. Universal or global indicators such as the CPI, the World Bank's Country Policy and Institutional Assessments (CPIA), and the World Bank Institute's Worldwide Governance Indicators (WGI) are not sufficient for this purpose.

#### *Establishing indicators and baselines*

In order to identify measurable and relevant indicators, ACAs should, first of all, *conduct stakeholder consultation* with donors, ACA staff, government agencies, experts, and civil society. This encourages common understanding of and buy-in for performance indicators. Stakeholder consultation improves the quality of the indicators and helps ensure that, at the end of the process, the evaluation findings will be considered credible by the various parties.

Second, it is necessary to *ensure that indicators are measurable*. An examination of available data sources should be undertaken to assess their relevance, applicability, validity, and reliability. Where possible, evaluators should not try to reinvent the wheel, but should use existing data sources. However, the ACA can add value by collating these data in a single database (more on electronic management information systems below). In general, the ACA staff responsible for M&E should work in collaboration with the country's statistics bureau (for general statistics), the judiciary (for crime statistics), and anti-corruption NGOs (for corruption perception surveys).

The process of indicator development is a core activity in establishing a good, results-based M&E system, as it drives all subsequent data collection, analysis, and reporting. There are several guidelines for indicator development that should be carefully consulted. None is specific to the area of anti-corruption, but their applicability is universal. A well-known mnemonic for indicator development is SMART. Indicators should be **S**pecific, **M**easurable, **A**ttainable, **R**elevant, and **T**ime-bound. Schiavo-Campo (1999) developed another set of principles with the acronym CREAM: indicators should be **C**lear (precise and unambiguous), **R**elevant (appropriate to the subject at hand), **E**conomic (produce data at a reasonable cost), **A**dequate (able to provide sufficient basis to assess performance), and **M**onitorable (amenable to independent validation). Both set of criteria are useful to remember and should be seen more as ideal types than as criteria for exclusion of indicators.

The right choice of indicators depends on the data systems, data collection methods, resources, and capacity of the ACA. As far as possible, the "sources of verification" and "comments/assumptions" columns in the indicator catalogue (annex 1) should note if a particular indicator needs special systems, methodologies, capacity, or other requirements. Finally, it is highly recommended that indicators be adapted and fine-tuned by the people who are going to use them, both before data collection begins and as an ongoing exercise. As suggested by Morra Imas and Rist (2009, 118), this can be done in a workshop where the following factors are listed for each indicator:

1. Data source
2. Data collection method
3. Who will collect the data
4. Frequency of data collection
5. Cost of collecting data
6. Difficulty of collecting data
7. Who will analyse and report data
8. Type of analysis (including limits and advantages of alternative approaches)
9. Who will use the data

The evaluations reviewed for the mapping exercise lacked well-developed indicators. Given that indicator development is critical to the whole process of evaluation, chapter 7 and annex 1 are devoted to this issue, providing examples and further guidelines. Once indicators have been identified and formulated, baseline data, benchmarks, and sets of cases for counterfactual scenarios need to be established.

It is important that donors base their evaluations on and contribute to the ACA's indicators, rather than imposing externally developed indicators on the ACA. There should be only one M&E system for the ACA, not one per donor. It is best practice to select only a few key indicators to measure success. Factors for choosing the indicators can be based on where public concern about corruption is focused (sectors, regions, or types of corruption) or where the greatest loss of state revenue is likely to be (central government versus local government, specific sectors, or specific institutions).

The ACA should conduct or commission surveys, qualitative evaluations, and risk assessments to establish baseline data for the identified indicators. These might include, for example, public perceptions of the work of the ACA or the number of corruption cases investigated per year. Case-processing time could be an example of an indicator for a process that the ACA wishes to improve and benchmark. Processing times for different types of cases could be mapped and performance improvements tracked periodically. Another useful way to use benchmarks would be to compare performances between ACAs with respect to how effectively and efficiently they accomplish their core functions. Because ACAs are so diverse, however (as described in section 4.1.2), cross-country comparisons often encounter a series of methodological difficulties.

#### 4.2.2 Building and running functional M&E systems

Evaluations depend on multiple sources of credible data. Many ACA evaluations are currently based only on key informant interviews, which limits their potential to (a) make strong, incontestable statements regarding performance and impact, (b) produce findings that can be generalised and used for comparative purposes, and (c) produce evidence to present to policy makers. A well-run operational M&E unit is often a goldmine for evaluators, particularly if indicators are established from the outset and if the M&E unit regularly collects data, supplementing its own record keeping with, for example, data from small surveys.

##### *The difference between outputs and outcomes*

The mark of a results-based M&E system is the ability to move from output level to outcome level. There is an unfortunate tendency amongst evaluators, donor staff, and ACA staff to take outputs as a proxy for outcomes. Therefore, it is important to make clear the critical difference between these two concepts. ACAs use inputs to carry out activities that result in outputs. Outputs should lead to desired outcomes. In the long term, outcomes should be reflected in broader results or impacts. Outputs are entirely under the control of the ACA, but outcomes are not. In other words, outputs can lead to predictable as well as unexpected outcomes. But, if the logic model or theory of change holds, we expect to see changes in outcome indicators. Examples of outputs include the amount of work done and the immediate results of activities. Outputs are products and services, and output indicators reflect changes in the supply of these products and services, such as the number of trainings held. Outcomes, in contrast, are the result of the use or application of outputs, such as, for example, public officials putting to use the knowledge they have acquired in anti-corruption trainings. Finally, outcomes reflect a programme's benefits, that is, changes in the situation, routines, perceptions, and so on, of the target group.

##### *Elements of an M&E system*

A key principle in building an M&E framework is to ensure that systems are in place, and responsibilities assigned, for collecting, monitoring, analysing, and storing data relevant to construction of the identified indicators. It is often necessary to employ a full-time person responsible for M&E, so that ongoing monitoring is established with less reliance on short-term consultants of varying quality; this also helps ensure institutional memory. This person should have a data collection and "compliance" function but should also work with senior management to ensure that plans, activities, and strategies are contributing towards the ACA's purposes and goals.

The M&E system should outline information requirements, data collection methods and intervals, and analytical frameworks, and establish an information management system to facilitate the storage, collation, analysis, use, and dissemination of data. The M&E system should include the logframe agreed with donors if one is in place; it needs to be practical and simple, yet capture all essential performance indicators. The M&E system should also clearly assign responsibilities for different M&E activities across the ACA departments.

An electronic management information system (MIS) for storing monitoring data should be an important component of the M&E system to ensure systematic data collection. Any proficient M&E system will result in the collection of large quantities of data and the development of significant amounts of information. In order for this information to be of both immediate and future value, data must be well managed, and this requires a sound storage and retrieval system. A simple database system based on Microsoft Excel or similar programmes could be developed. More advanced systems with ability to cross-reference and use data from geographic information systems could also be options if resources are sufficient and spatial data are important. The important point is that the MIS should be able to capture and retrieve data in a way that facilitates the sharing, reporting, use, and dissemination of information.

It is important that ACA staff understand that honesty is the most essential part of any M&E system and do not attempt to hide or suppress errors or failures. It is also important for ACA senior management to create a culture of learning in which the honest reporting of successful and unsuccessful outcomes is supported. Lessons can only be learned and applied if the ACA's work is systematically reported and analysed and the analyses are fed into future decision-making processes.

There are several core guidelines for developing an ACA M&E system:

- Set priorities. Focus on a few indicators at the outcome and impact levels, particularly if data collection external to the ACA is also required.
- Develop the reflective and analytical capabilities of those involved, including senior management and programme officers as well as the M&E officer.
- Place emphasis on analysis and decision making.
- Be sure to capture negative and unintended change, in addition to expected or planned change.
- Structure the system around learning.
- Produce information that is useful for all levels of stakeholders.
- Issue thorough and regular monitoring reports based on consistent tracking of selected indicators.

Successful development, implementation, and sustainability of the M&E system require the following essential building blocks, building on the above principles:

- *Vision:* An understanding of how M&E information can support management and decision making. This requires strategic leadership as well as a clear understanding of the basic concepts and potential uses of the system.
- *Enabling environment:* A commitment not only to launching the M&E system but also to sustaining it over the long term.
- *Technical capacity and infrastructure to supply M&E information:* This includes credible and relevant data- and information-gathering systems (such as small surveys), as well as skilled personnel to gather, analyse, and report on progress towards objectives. Putting in place a data system may require training M&E specialists and the technical staff to effectively use the system. This is a lengthy and iterative process; continuous learning will be beneficial to the improvement of the M&E system.
- *Infrastructure to demand and use M&E information:* This requires both informal mechanisms (e.g., opportunities for reflection) and formal ones (e.g., an integrated MIS system) for reporting and sharing M&E information.

In short, M&E systems for ACAs do not have to be complicated. They can be based on simple, creative processes which utilise participatory stakeholder feedback, sound analysis, and multiple tools to encourage active stakeholder engagement. The resources and time required to develop such a system would be well spent.

### *Managing evaluation processes*

ACA staff in the M&E unit should have the responsibility, potentially together with donor counterparts, for managing the evaluation process. Several conditions will facilitate a productive evaluation. First, expectations for an ACA should be clearly outlined from the outset in criteria for its success, and it should be established whether the ACA has the required resources and mandate to meet these criteria. ACAs need to make sure that a cause-consequence logic, or a theory of change, can be established. Second, efforts should be aimed primarily at measuring outcomes rather than outputs and activities, which show little about impact. Third, it is important not to overload the evaluation by including too many indicators. Fourth, ACAs should remember that evaluations benefit from mixed methodologies and should use both quantitative and qualitative indicators. Finally, evaluations should be made available for public scrutiny. Getting feedback from citizens, donors, and other stakeholders is important for improving the validity of selected indicators, the focus of the analysis, the consideration of counterfactual scenarios, and ultimately the quality of the work of ACAs.

### *Promoting learning*

The M&E system should be geared towards learning, not just accountability. An approach to M&E limited to accountability reflects an audit mind-set and often works mainly with checklists. It reviews whether the ACA has spent its resources (inputs) on the planned activities and developed the expected products (outputs). A learning-oriented approach gives priority to outcomes and impacts. It tries to use data to explain why performance was either good or bad and to come up with suggestions for improvement. Chapter 8 provides more information on the importance of learning in M&E, along with guidance on how ACAs can adopt a systematised approach to learning in their M&E frameworks.

#### **Box 4. Elements of a good M&E system**

- Clear statements of *measurable objectives*
- A structured set of *indicators* clearly linked to the measurable objectives
- Provisions for collecting and managing *project records* so that the data required for indicators are available, at reasonable cost
- Arrangements for *collecting, analysing, and reporting data*
- Arrangements to ensure that results and recommendations *influence decision making*

## 5. How can we perform better evaluations?

This chapter introduces best-practice principles for evaluation, principles which have been used in other sectors for decades. Adopting these principles should facilitate changes in the current practices of ACA staff, donors, and evaluators, in which evaluations are mainly conducted as box-ticking exercises. The aim is to go beyond paying lip service to the concepts of “outcomes” and “impacts” and invest the work and resources needed to evaluate them meaningfully.

As shown in chapter 2, evaluations of ACAs generally fall short of establishing evidence-based findings on the outcomes and impacts of ACAs. This is because they tend to focus exclusively on the ACA’s organisational performance at the levels of inputs, activities, and outputs. Thus an ACA can be seen as performing well but may still fail to have meaningful impacts or bring about sustainable positive change. This chapter outlines how the outcomes and impacts of ACAs can be better evaluated, so that ACAs can improve the quality of their work and respond to external pressures for results.

For evaluations in areas that require “complex evaluations,” such as anti-corruption, human rights, women’s empowerment, and so on, the point of departure should always be to look back at the problem that the intervention, in this case the creation of an ACA, was supposed to solve. Change is then measured in relation to that baseline negative situation. This requires acknowledging that “light” evaluation methodologies are not sufficient and that complex evaluation strategies are needed.<sup>12</sup>

The chapter builds heavily on existing OECD evaluation standards, but it is tailored for ACA purposes in particular. For experienced evaluators, some repetition is unfortunately inevitable. We have chosen to restate what might be obvious for some people in order to ensure that all audiences speak the same language and adhere to the same standards.

Evaluators and ACA staff are the primary target groups for this chapter. However, donors can also benefit from the guidelines on how good evaluations can be designed and performed, *inter alia* in drafting the terms of reference (ToR) for an evaluation and in obtaining quality assurance.

Seven issues are of particular importance for the quality of an evaluation of ACAs:

- Purpose and rationale for evaluation
- Composition of the evaluation team
- Evaluation scope, overall principles, methodology, and questions
- Analytical methods for the evaluation
- Participation and transparency
- Quality assurance
- Budget and financial aspects of M&E

---

<sup>12</sup> We base our definition of a complex evaluation on the one provided by EuropeAid, namely, an evaluation that involves multiple activities, heterogeneous target groups, many expected effects, and several expected impacts (European Commission 2006b, 17–18).

## 5.1 Purpose and rationale for evaluation

The first important step in any evaluation is to agree on its focus and purpose. This should be clearly communicated to all stakeholders and to staff at all levels within the ACA; it should also be written in the evaluation ToR. The latter should explicitly state the objectives of the evaluation, how findings will be used, the required level of precision, and the kind of policy and/or management decisions the findings should inform.

It is also important to describe the background for the evaluation and who initiated it. Current practice is that evaluations on the performance of ACAs are initiated either by the institution itself, by the government, or by the funding/supporting donors.

The evaluation objectives should later be translated into relevant and specific evaluation questions (see section 5.3).

### **Box 5. Specifying the rationale, purpose, and objectives of an evaluation**

According to the OECD/DAC's *Quality Standards for Development Evaluation*:

The rationale, purpose and intended use of the evaluation [should be] stated clearly, addressing: why the evaluation is being undertaken at this particular point in time, why and for whom it is undertaken, and how the evaluation is to be used for learning and/or accountability functions.

For example the evaluation's overall purpose may be to:

- contribute to improving a development policy, procedure or technique,
- consider the continuation or discontinuation of a project or programme, or
- account for public expenditures and development results to stakeholders and taxpayers.

The specific objectives of the evaluation clarify what the evaluation aims to find out. For example to:

- ascertain results (output, outcome, impact) and assess the effectiveness, efficiency, relevance and sustainability of a specific development intervention, [or]
- provide findings, conclusions and recommendations with respect to a specific development intervention in order to draw lessons for future design and implementation.

*Source: OECD/DAC 2010b.*

## 5.2 Composition of the evaluation team

The complexity of ACA evaluations may require the participation both of experts in evaluation techniques and of specialists in the area the organisation works on, such as asset declarations, conflicts of interest, etc. Factors in choosing evaluators include credibility, expertise, and personal distance from the organisation under review. However, members of the ACA in question should also take part

in the evaluation, as this is a useful way of gathering an inside understanding of organisational processes and promoting buy-in of the evaluation results.

The balance of outside and inside expertise is likely to be more problematic in cases where donors request the evaluation. These often involve a team of international experts who are “parachuted in.” In our experience, such quick evaluations, which often rely on the expert opinion of one or two individuals, rarely produce useful information. Instead, donors would be well advised to support ACAs in improving their own M&E systems and self-assessment capacity. They should also explore the possibility that local or regional experts could conduct the evaluation without compromising its quality.

It is important that evaluators be independent from the organisation under evaluation, including its policy, operations, and management functions, as well as from intended beneficiaries. Possible conflicts of interest should be addressed openly and honestly. The fact that the majority of the evaluations reviewed in chapter 2 did not discuss these issues is a grave concern.

## 5.3 Evaluation scope, overall principles, methodology, and questions

### 5.3.1 Scope of the evaluation

It is important to clarify the ambition, working areas, and resources of the evaluation at the outset. This requires that the ACA be able to clearly define its vision and mission and describe its theory of change. Section 4.2.1 provided detailed guidance on how theories of change for ACAs can be developed. The evaluation scope defines the time period, funds spent, geographic area, target groups, organisational set-up, implementation arrangements, policy and institutional context, and other dimensions to be covered by the evaluation (OECD 2010a, 11).

It is also important to clarify whether the evaluator is being asked to do a rapid assessment, an organisational capacity assessment, or an evaluation. If the OECD evaluation criteria of relevance, effectiveness, efficiency, impact, and sustainability are used (see below), will they all be covered? The greater the scope of the exercise, the greater the resources required.

### 5.3.2 Evaluation principles and methodology

Evaluators need principles and standards for their profession that promote accountability, facilitate comparability, and augment the quality and reliability of their products (Picciotto 2005, 30–59). As a minimum, the methodology for the evaluation should clarify the following aspects of the evaluation’s approach:

- Choice of evaluation area
- Formulation of evaluation questions
- Choice of methods
- Formulation of research design and data collection techniques
- Implementation of data collection
- Analysis of data
- Interpretation of data
- Conclusions<sup>13</sup>

---

<sup>13</sup> Bryman 2004, 21–22.



The literature on how to conduct good evaluations in the area of development assistance is substantial and well developed. It is therefore beyond the scope of this report to give a full overview. However, our review of the evaluations of ACAs revealed that none adhered fully to established guidelines. Therefore, we present a brief section on good evaluation guidelines, focusing on the OECD/DAC Principles for Evaluation of Development Assistance from 1991 (OECD/DAC 1986, 1991, 2000).

The evaluation criteria recommended by the OECD are relevance, effectiveness, efficiency, impact, and sustainability. These are normally the backbone of evaluation ToRs. The ACA evaluations focused mainly on effectiveness, and to some extent on efficiency. Some, when assessing the mandate and functions of the ACA, approached the subject of relevance. However, none fully met the OECD criteria noted in box 7.

#### **Box 6. Research design**

A research design provides a framework for the collection and analysis of data. A choice of research design reflects decisions about the priority given to a range of dimensions of the research process. These include the importance attached to (a) expressing causal connections between variables; (b) generalising to larger groups of individuals than those actually forming part of the study; (c) understanding behaviour and the meaning of that behaviour in its specific context; and (d) having a temporal (i.e., over time) appreciation of social phenomena and their interconnections.

*Source: Bryman 2004, 27.*

All five OECD criteria need not be within the scope of the evaluation. It would normally require too many resources to address them all, and the result might be that the evaluator would not succeed in covering any of them very well. Effectiveness and impact are generally considered most important, but it all depends on what specific problems are to be evaluated. (See chapter 6 for more on how to undertake impact evaluations.) If an ACA is thought to be wasting its resources, compared to other ACAs, then efficiency would be the key principle to investigate. If, on the other hand, it is the relevance of the ACA in the overall institutional setting of the country that is questioned, then relevance is what the evaluator should focus on. Measuring effectiveness and impact also generally requires data time-series, so early evaluations might have to focus on efficiency until additional data become available.

Evaluation questions should be determined early on in the process and should inform the development of the methodology. According to the OECD,

A methodology includes specification and justification of the design of the evaluation and the techniques for data collection and analysis. The selected methodology provides for answering the evaluation questions using credible evidence. A clear distinction is made between the different result levels, with an intervention logic containing an objective-means hierarchy stating input, output, outcome, and impact. (OECD 2010a, 24)

Development of a methodology includes choosing analytical methods, as described in section 5.4.

**Box 7. OECD/DAC criteria for evaluating development assistance**

*Relevance:* The extent to which the aid activity is suited to the priorities and policies of the target group, recipient, and donor. In evaluating the relevance of a programme or a project, it is useful to consider the following questions:

- To what extent are the objectives of the programme still valid?
- Are the activities and outputs of the programme consistent with the overall goal and the attainment of its objectives?
- Are the activities and outputs of the programme consistent with the intended impacts and outcomes?

*Effectiveness:* A measure of the extent to which an aid activity attains its objectives. In evaluating the effectiveness of a programme or a project, it is useful to consider the following questions:

- To what extent were the objectives planned (goals and outcomes) achieved [or] are likely to be achieved?
- What were the major factors influencing the achievement or non-achievement of the objectives?

*Efficiency:* Efficiency measures the outputs—qualitative and quantitative—in relation to the inputs. It is an economic term which signifies that the aid uses the least costly resources possible in order to achieve the desired results. This generally requires comparing alternative approaches to achieving the same outputs, to see whether the most efficient process has been adopted. When evaluating the efficiency of a programme or a project, it is useful to consider the following questions:

- Were activities cost-efficient?
- Were outputs achieved on time?
- Was the programme or project implemented in the most efficient way compared to alternatives?

*Impact:* The positive and negative changes produced by a development intervention, directly or indirectly, intended or unintended. This involves the main impacts and effects resulting from the intervention on the local social, economic, environmental, and other development indicators. The examination should be concerned with both intended and unintended results and must also include the positive and negative impact of external factors, such as changes in attitudes towards corruption amongst the general population. When evaluating the impact of a programme or a project, it is useful to consider the following questions:

- What has happened as a result of the programme or project?
- What real difference has the activity made to the beneficiaries?

*Sustainability:* Sustainability is concerned with measuring whether the benefits of an activity are likely to continue after donor funding has been withdrawn. When evaluating the sustainability of a programme or a project, it is useful to consider the following questions:

- To what extent did the benefits of a programme or project continue after donor funding ceased?
- What were the major factors which influenced the achievement or non-achievement of sustainability of the programme or project?

*Source: Adapted from OECD/DAC 2011.*

## 5.4 Analytical methods for evaluation

A range of analytical methods can be used to determine outcomes and impacts for ACA interventions. As mentioned above, current ACA evaluations often make the mistake of taking organisational capacity assessment to be the best methodology for assessing the performance of an ACA, and they often rely on interviews as the sole analytical method. The fact that ACAs are organisations does not mean that one can't measure the outcomes and impacts of their specific interventions using other methods.

The choice of analytical methods depends on the evaluation questions. If the focus of the evaluation is on efficiency, then the evaluator will be concerned with identifying the ACA's activities and measuring the internal resources used and outputs produced. This need not entail advanced analytical techniques. However, if the focus is on effectiveness or impact, then the evaluator will need to measure outcomes and impacts. This often calls for more sophisticated analytical techniques.

The presentation below is based on the work of McGee and Gaventa (2010). These authors identify several studies which have used analytical methods that can show evidence for either contribution or attribution of impact.<sup>14</sup> They discuss four methods that have been used successfully:

- *Surveys*: These can be used for a multitude of purposes, and stakeholders often like the “hard” data obtained through surveys. As an example of a quantitative method, the Public Expenditure Tracking Survey (PETS) took a “follow the money” approach, comparing resources received with resources actually spent to account for leakage. Examples of surveys which focus not on discrepancies but on perceptions are Quantitative Service Delivery Surveys (QSDS) and Citizen Report Cards (Reinikka and Svensson 2000, 2003; Sundet 2008; Ravindra 2004). While ACAs can learn from these sophisticated methodologies, their own surveys do not have to be complex or comprehensive to allow for outcome and impact measurements. Quantitative surveys can allow for advanced statistical analysis (such as multivariate regression analysis or difference-in-difference methodology, which is used for impact evaluations), if the data are comprehensive enough. But more qualitative surveys also have the potential to be very useful for measuring ACA performance.
- *Experimental approaches*: These include randomised controlled trials, which uphold basic scientific principles for constructing evidence. Such trials have been used in the field of medicine for decades and have also been used more recently in economics. Use of experimental approaches in the social sciences is not often possible, and some critics would say it is not even advisable. However, evidence produced using experimental or quasi-experimental designs typically carries greater weight than that from studies which do not use these designs. A quasi-experimental study is one which fulfils certain characteristics of experimental designs, but that does not fulfil all of the internal validity requirements (Bryman 2004, 39). Examples of relevant evaluations using experimental approaches are those by Olken (2007) and Bjorkman and Svensson (2009). Olken's study used a basic experimental design to examine the difference in effect of external audits versus community oversight and grassroots monitoring in preventing corruption in 600 village road projects. This large sample allowed randomisation, a core feature of an experimental design. Essentially, through a random allocation, some projects were chosen to be audited and others to have community oversight mechanisms. Thus, Olken was able to incorporate the evaluation design into the programme design at an early stage;

---

<sup>14</sup> A presentation of analytical methods suitable for more advanced readers is also given by Nick Duncan (2006).

otherwise an experimental approach would not have been possible. Olken also used stratification as a method to avoid contamination or spillover effects.

- *Qualitative case studies:* Case study analysis is often used in the field of corruption. Indeed, when there is no explicit methodology, evaluations are often simply labelled as “case studies.” However, case studies also have methodological guidelines which need to be upheld; see, for example, Yin (2003) and Stake (1995). Case studies go well together with other methodologies as a way to contextualise findings and provide possible cause-effect explanations. Wade (1982, 1985) produced early examples of informative case studies.
- *Interviews:* This is a popular and widely used qualitative method which is often combined with other methods (both qualitative and quantitative) in a systematic fashion to yield a combined research methodology. Interviews can be structured, semi-structured, or unstructured. If several people are interviewed on a specific topic, a focus group method can be used.

In general, triangulation is recommended as a research strategy. Relying on a single data source or analytical technique in a complex field such as anti-corruption reduces the confidence one can have in the findings of an evaluation. Triangulation is a way to cross-check the credibility of the conclusions. Good research strategies often include both qualitative and quantitative methods (mixed methods), used sequentially in an iterative process. An iterative research strategy promotes a repetitive interplay between the collection and analysis of data; in other words, analysis begins after some of the data have been collected, with the preliminary findings then shaping the next step in the data collection process (Bryman 2004, 399).

## 5.5 Participation

Too often, evaluations are conducted as an external exercise, without significant involvement of stakeholders and beneficiaries. This also seems to be the case with the ACA evaluations reviewed in chapter 2. Participation should be promoted, not as an end in itself, but because it enhances the quality of evaluation by providing validation of data and findings. It also has significant implications for long-term sustainability, since recipients are solely responsible after the donor has left.

The terms of reference for an evaluation should be drafted so that, whenever possible, ACA staff, donors, and external stakeholders are involved in the evaluation process. External stakeholders could be NGOs working on accountability issues or government agencies involved in the fight against corruption. Such stakeholders also often have useful data to contribute. Beneficiaries—the population in general and people who have benefited from the ACA’s services in particular—should also be surveyed to elicit their views. Too often, ACA evaluations are left to “expert judgements.” But the views and expertise of affected groups should also be an integral part of the evaluation. They should be involved at an early stage in the evaluation process and given an opportunity to contribute to evaluation design by, for example, identifying issues to be addressed and evaluation questions to be answered (OECD 2010a, 9, 23).

Stakeholders can be defined as people or organisations that are likely to be affected by the work of the ACA and/or can influence the success or failure of its work. Various tools for stakeholder identification exist, such as constructing Venn or Spider diagrams.<sup>15</sup>

---

<sup>15</sup> For further guidance on how to conduct stakeholder analysis, the five-step approach described in the EuropeAid Project Cycle Management Guidelines (European Commission 2004) is useful.

Transparency should also be promoted. Evaluations should be made available for public scrutiny. Feedback from citizens, donors, and other stakeholders can improve the validity and reliability of the conclusions reached.

## 5.6 Quality assurance

This section is relevant for ACA and donor staffs that manage ACA evaluations. As noted above, the OECD/DAC Network on Development Evaluation established the “DAC Evaluation Quality Standards” in 2006, setting out 10 criteria for assessing evaluation quality (OECD/DAC 2006; see also OECD/DAC 2010b). These criteria were used for the mapping exercise in chapter 2: rationale, purpose, and objectives of an evaluation; evaluation scope; context; evaluation methodology; information sources; independence; evaluation ethics; quality assurance; relevance of the evaluation results; and completeness.

In the future, ACA and donor staff should hold evaluators accountable for at least these criteria/quality assurance standards and ensure that their ToRs reflect this. Quality assurance begins with well-constructed ToRs. During the evaluation, it is important that the ACA and/or donor evaluation manager supervise the process in a constructive manner. Once a draft version of the report has been written, a review process is an integral part of quality assurance. The evaluation plan should allow adequate time for circulating the report, for receiving comments on it, and for incorporating valid comments and qualifications into the final version. Any unresolved differences of opinion should be clearly stated in the report. Peer review is also a good method for quality assurance. ACAs should share evaluation findings, and evaluators are encouraged to consult other ACAs if they are in doubt.

## 5.7 Budget and financial aspects of M&E

As shown in chapter 2, the budgets for ACA evaluations are often insufficient to cover a thorough evaluation. Donors need to be more realistic about what good evaluations cost, but more importantly, they should use their resources better to achieve more value for money.

Currently, ACAs are assessed frequently but only with “light” evaluations. More value for expenditure would be obtained if money were spent on designing an M&E system, with ACA staff, that can regularly collect monitoring information on outputs, outcomes, and impact indicators. The cost of building and maintaining the M&E system should be included in the regular budget; for example, around 5–10 per cent of the total budget could be set aside for M&E. As this would also pave the way for better results-based management, and results in capacity building, the money would be well spent.

As emphasised above, the key principle for building an M&E system is to keep it simple, and to identify indicators and small surveys for outcome and impact measurements. With M&E systems in place, donors should be able to aim for bigger external evaluations every five to seven years. This might free up funds to do so-called “impact evaluations,” which are explained in chapter 6.

## 6. Impact evaluation methodology

ACAs are under pressure from donors to show impact. National governments and the public are also increasingly demanding results. But ACAs do not currently have the tools to respond to these pressures by providing evidence to show their value to society.<sup>16</sup> Impact evaluations, important elements in evidence-based policy making, can provide the necessary information base for such communication strategies.

Impact evaluations are receiving increasing attention, particularly from donors, because they can take evaluation beyond the output level. To date, despite enormous amounts of funds spent on implementing development programs, donors and their local partners have little evidence of the impact of their financial support. This is partly because insufficient attention and funding have been devoted to assessing impact, but it also reflects the mixed quality of impact assessments to date. This knowledge gap has been well documented for the social sectors in the report *When Will We Ever Learn?* from the Center for Global Development (2006).

Impact evaluation methodology has not yet been applied in the area of anti-corruption.<sup>17</sup> However, this does not mean it cannot be done. Anti-corruption interventions are not more difficult to measure than those in many other sectors where impact evaluations are currently being undertaken. This chapter does not present a complete impact evaluation design, as there are many different models one could use, depending on the ACA in question and on how ambitious an evaluation is required. However, the chapter does provide guidance on the basic elements required for an ACA impact evaluation.

Many evaluations claim to measure impact, but few use a rigorous impact evaluation methodology that actually account for impact. We do not expect all evaluations to be full-scale impact evaluations, but the principles behind impact evaluations are still useful for any kind of evaluation as guidance for strengthening the research design.

Bamberger (2006, 2–3) establishes three minimum requirements for quality impact evaluations:

- Develop a *set of indicators* that can meaningfully and reliably define and measure project inputs, implementation processes, outputs, outcomes, and impacts.
- Develop a logically sound *counterfactual* statement presenting a plausible argument that observed changes in outcome indicators after the project intervention are in fact due to the project and not to other unrelated factors, such as improvements in the local economy or programs organized by other agencies.
- Determine, in accordance with accepted statistical procedures, *whether a project has contributed to the intended impacts* and benefited a significant proportion of the target population.

The methodological approach to impact evaluation taken in this chapter is pragmatic. We share the belief of several authors that “the most appropriate methods should be identified to fit the operational

---

<sup>16</sup> A recent United Nations Office on Drugs and Crime / World Bank publication examines the importance of public support for ACAs and outlines the communication tools they need to build such support. See Byrne, Arnold, and Nagano (2010).

<sup>17</sup> There are a multitude of macro-level surveys with “global” indicators that seek to measure changes at the impact level. While they provide valuable data, they are not evaluations as such, since they do not attempt to show the results of particular interventions.

context, and not the other way around” (Gertler et al. 2011, xiii–xiv). This is best done through a so-called prospective approach which builds the evaluation design into the programme implementation design.

## 6.1 Elements of an ACA impact evaluation model

Evaluations of ACAs will not always be able to use the statistically strongest research design, based on randomisation or statistical matching. This depends on the possibility of doing prospective evaluation and on the quality of data in the individual case. In most cases it can be expected that quasi-experimental or non-experimental designs will have to be used. However, this weakness can be counterbalanced by strengthening the evaluation design in other methodological dimensions.

Scholars differ on the question of whether a weak statistical design can be fully compensated by strengths in other methodological dimensions, allowing the result to be called an “impact evaluation.” One group of scholars, labelled “randomistas,” argue that randomized experimental design should be the gold standard for impact assessment, whereas others say that strong quasi-experimental designs provide an acceptable next-best option. Evaluation designs with counterfactual statements supported by analysis of statistically matched comparison groups are also considered a requirement for a “proper” impact evaluation by some authors. Others are willing to accept the use of alternative bases for counterfactual arguments, such as secondary data, programme theory models (theories of change), outcome mapping, or concept mapping. For most anti-corruption practitioners, however, this is largely a theoretical discussion. In practical terms, the issue is how to implement the best possible approach to measuring impact of anti-corruption projects that is actually feasible to execute.

### Box 8. Criteria for evaluation of social research

Three of the most important criteria for the evaluation of social research are reliability, replication, and validity:

- *Reliability* is concerned with the question of whether the results of a study are repeatable, and whether measures are consistent.
- *Replication* is closely associated with reliability. If the researcher spells out his or her methods, there is a greater chance that the study will be replicable.
- *Validity* is concerned with the integrity of the conclusions that are generated from the research. This can be broken down into (a) measurement validity, (b) internal validity, (c) external validity, and (d) ecological validity.

Source: Bryman 2004, 28.

The authors of this manual advocate three methods in particular to build the strongest possible evaluation design of ACAs, regardless of whether randomisation or statistical matching is possible:

- Basing the evaluation on a strong *theory of change*, deriving indicators and research hypotheses and questions from this;
- Using an evaluation design with *mixed methods*; and
- Utilizing *multiple indicators of different types at both the outcome and impact levels* to increase construct validity, that is, breaking down complex constructs such as corruption with more detailed analysis (Bamberger, Rugh, and Mabry 2006; Bamberger and White 2007).

These three points are discussed in more detail in the subsections below.

### 6.1.1 Theory of change

The basic features of the theory of change logic have already been covered in the chapters above. Theory-based evaluation, that is, “examining the assumptions underlying the causal chain from inputs to outcomes and impact,” is a well-established approach advocated by, *inter alia*, White (2009). Using a theory-based approach to impact evaluation means that we focus on understanding not only *what* works, but *why* it works, even when randomisation and experimentation are difficult to include in the evaluation design. Theory-based impact evaluation is thus suitable for measuring the impact of an intervention such as an ACA, where it is often not possible to have a control or comparison group.<sup>18</sup>

In a comparison of three different types of common interventions, Ravallion (2009) explored the potential for evaluating the impact of anti-corruption commissions and recommended a theory-based approach. The options he proposed included phasing in geographic or ministerial coverage, so that the impact of an ACA intervention (such as a whistleblowing scheme) on a certain ministry or local police force could be measured against cases where no such intervention was made. These cases could then serve as counterfactual scenarios.<sup>19</sup>

As most anti-corruption practitioners know, an issue for impact evaluations is that theoretical explanations of corruption differ greatly. Therefore it is essential to reach agreement among a wide range of stakeholders on the programme’s theories of change. This should not necessarily be difficult, since theories of change are not established truths but rather crystallisations of the implicit logic of the programme. Every programme has a logic underlying its interventions and it is this logic of causal pathways which should be tested.

A simple example of a causal chain could be that donor funding (input) enables the ACA to implement an online complaints system (activities), leading to  $x$  number of investigations (output), which if done right can lead to  $x$  number of convictions (outcomes), and in time also to reduced corruption levels (impact). This is, of course, a simplified model; a full causal chain would also include many other contributing elements.

### 6.1.2 Mixed methods

A mixed methods design is one in which both quantitative and qualitative methods of analysis are used in an iterative and complementary way. Each method has its strengths and weaknesses; by combining them, one achieves a better evaluation result.

Mixed methods design can address many typical problems affecting the validity of statistical impact evaluations. Some of the main advantages include:

- *More precise measurement, increasing construct validity:* As with using multiple indicators, the use of mixed methods design can increase the construct validity of an evaluation. Construct validity is extent to which an evaluation actually measures the variable it is intended to measure. (A classic question of construct validity is to what extent an IQ test actually measures intelligence.) Corruption is a complex social construct, and it is important that evaluations actually measure the intended aspects of corruption.

---

<sup>18</sup> ACAs are, in the words of Martin Ravallion (2009, 227–36), not typically an “assigned programme”: that is, they do not normally only target certain observational units, but have nationwide targeting. This makes it more difficult to establish a control group.

<sup>19</sup> A certain selection bias would potentially hamper fair evaluation, as units of treatment would most likely be selected because of higher prevalence of corruption and therefore would differ from the control group. This is, however, not an insurmountable problem and can be dealt with in a variety of ways.



The study of corruption often relies on proxy variables that do not fully cover the construct (corruption) which one seeks to analyse. Using qualitative methods, such as exploratory case studies, focus groups, or interviews, the evaluator can strengthen the understanding of the key variable, namely corruption.

- *Contextualising the analysis:* Purely quantitative statistical impact evaluation designs may have strong internal validity. But they may fail to take into account political, institutional, historical, and sociocultural factors that can make interventions fail or succeed. Building political economy and contextual analysis into the evaluation design will make the findings more credible, as it will result in a stronger understanding of the theory of change. Understanding the social, political, and economic setting in which an intervention takes place is key to identifying its impact. For example, the implementation of an online complaints system will have to take into consideration not only developmental issues (e.g., levels of computer literacy and access to the Internet) but also technical issues (e.g., whether to secure the user's identity) and cultural issues (e.g., tendency not to report due to fear of reprisals). In addition, legal issues (e.g., the existence and enforcement of whistleblower protection legislation) will have an external effect on the level of complaints.
- *Addressing the problem of the “black box” and the process of project implementation:* Most quantitative impact evaluations focus only on whether the project has achieved its objectives by comparing pre-intervention levels with post-intervention levels. This makes it impossible to determine whether poor results are due to design failure or implementation failure. Qualitative techniques can capture project processes and triangulate this with the quantitative data.
- *Extensive use of triangulation:* Triangulation entails using more than one method or source of data in the study of social phenomena, as originally conceptualised by Webb et al. (1966). Denzin (1970, 310) uses the terms more broadly to refer to an approach that combines “multiple observers, theoretical perspectives, sources of data, and methodologies.” By using both quantitative and qualitative methods, the evaluators can assess the validity of different data sources by systematically comparing them with each other and conducting consistency checks (Bamberger, Rao, and Woolcock 2010).

In sum, including mixed methods in the methodology is highly recommended for evaluations of ACA. So far, qualitative approaches (interviews, desk reviews, etc.) have tended to dominate ACA evaluations. This is partly because ACAs have not developed results-based M&E systems for baseline data and systematic data collection. ACAs need to devote more time and resources to establishing quantitative approaches, and then combine them with qualitative methods.

### 6.1.3 Multi-level, multi-type indicators

Anti-corruption practitioners are particularly prone to relying on a single indicator or index to account for impact. This is typically one of the macro-level indexes such as Transparency International's CPI or the World Bank's WGI. This may be a result of the so-called “data revolution,” which popularized these indexes in the 1990s. However, relying on just one indicator or index for a complex phenomenon such as corruption is a faulty approach. One cannot infer that one single institution has a causal effect by measuring a macro-level outcome. As noted above, indicators should be disaggregated as far as possible to obtain the most specific measure of the construct possible.

However, there will always be a call for impact-level measurements at a general level. The fact that many of the corruption indexes are not always consistent with each other is a problem, one that arises in part from the many validity problems inherent in their methodologies. Several scholars have

highlighted the pitfalls of using such indexes, but their popularity has not decreased (Galtung 2005; Søreide 2003, 2006; Arndt and Oman 2006; Knack 2006).

Whenever resources permit, it is also advisable to triangulate by complementing generic macro-level indexes with indicators specifically developed for the purposes of the evaluation, as it is rare that generic indicators fully capture the programme logic. Perception-based indicators would ideally be combined with “objective” indicators based on facts or experience (see chapter 7). Indicators should also reflect the fact that positive outcomes and impacts can include not only reduced corruption levels and measures of crime statistics, but also improvements in relating to partner organisations, public awareness, skills upgrading, etc. An ACA with no investigatory or prosecutorial powers should not be expected to contribute to changes in crime statistics, for example. Even ACAs with investigatory powers are often not concerned with quantitative measures of impact, since major complicated corruption cases may be so prominent that numerical measures such as numbers of cases prosecuted are irrelevant. Chapter 7 is devoted to the issue of indicators and their importance for monitoring and evaluation.

#### 6.1.4 Summary

One should always strive to use the best evaluation design possible. However, in the real world there are always obstacles and limitations, and evaluators often have to improvise, adapt, and overcome. Even if a statistical experimental design is not possible, impact evaluations of ACA can be based on quasi-experimental or non-experimental designs as long as they are supported by at least three methodological pillars: (a) a strong theory of change design, (b) a mixed methods design, and (c) the use of multi-level, multi-type indicators.

In short, even if there are methodological challenges which make reaching the highest standards unfeasible, there is no reason why donors should not promote impact evaluations in the area of anti-corruption, particularly of ACAs. Statistical shortcomings can be counterbalanced by strengthening the evaluation design on other fronts.

The rest of this chapter comments on specific methodological issues relating to impact evaluations that are of interest to experienced evaluators. It first discusses current practices in impact assessment and then considers how to deal with the attribution issue. Finally, options for constructing counterfactual scenarios for ACA interventions are presented.

## 6.2 Current practices in measuring ACA impact

Donors and evaluators currently measure impact in the area of anti-corruption in two principal ways. Both are deceptive. First, some evaluations are called “impact assessments” even though they do not actually evaluate impact. Second, “global indicators,” cross-country indexes, international rankings, and the like are sometimes used to make claims about impact when such a connection is not verified.

Regarding the first issue, an evaluation may be tasked with impact measurement even though it does not have an evaluation design capable of such a measurement. Donors should not yield to the temptation to commission so-called “impact assessments” unless they are willing to provide evaluators with the comparatively substantial resources required for such an evaluation (rather than only those needed for a more modest light organisational assessment). Donors should not easily accept assurances from evaluators that they can carry out such an assignment without adequate resources. None of the evaluations reviewed in chapter 2, for example, qualifies as an impact assessment.

Another bad practice that became clear from the mapping exercise is that donors typically set a time horizon for evaluation that is too short. A time lag will always exist from the point the intervention begins until one can reasonably expect to see any change in outcomes, let alone impact. Thus, an impact evaluation of an ACA after two years of its existence would not be sufficient to determine its

success. It could only serve as a first indication of its performance. This does not mean that one should not attempt to measure impact. But such evaluations should be done rigorously, should be based on evidence and an understanding of the mandate, functions, and context of the ACA, and should take into account the appropriate time frame.

The second issue, that of global indicators, is more complex and is related to the discussion about accountability and attribution in section 6.3. Global indicators, cross-country surveys, and international indexes are easily accessible and informational. The questions they ask do refer to the impact level (such as the perceived levels of corruption in a country). However, that does not mean that these data can be used to measure the degree to which an ACA has succeeded in meeting its goals, even if those measures do correspond to goals such as reducing the levels and scope of corruption or changing values in a society. The problem is that “it is almost impossible to identify causal links between ACAs and these macro-level outcomes, especially in light of the need to account for the influence of structural reforms and other important factors. Thus, we should be wary of simplistic assertions about any ACA’s impact on corruption. Indeed, many agencies’ missions are broadly defined in terms of reducing corruption or changing values—outcomes that are, at best, very hard to measure” (USAID 2006, 4). If one only uses generic global indicators (combined with some output-level measurements) to measure ACA performance, it is indeed impossible to identify causal links. However, if the methodology proposed in this chapter is used (using theory of change, mixed methods, and multi-level, multi-type indicators), causal links can be plausibly identified, and impacts as well as outcomes can be measured.

In short, one cannot infer impact of a specific project or institution from changes in general indicators. In order to assess the impact of the work of an ACA, the evaluation must be approached from the perspective of the particular programme being implemented. Global indicators, in sum, cannot be used as shortcuts to measure ACA impact.

### 6.3 Issues of accountability and attribution

A solid understanding of impact evaluation methodology can contribute to closing or readjusting the current gaps between expectations and reality which exist for some ACAs. Expectations are often high for ACAs, but they are not always given the resources, political space, or time necessary to live up to these expectations. Expectations of direct causal effects from ACA interventions at the outcome and impact levels may be unrealistic. In most areas of anti-corruption, researchers have not yet even solved overall causality problems. We understand that there is a correlation between transparency and lower levels of corruption. But we do not know with certainty whether transparency leads to less corruption, or whether transparency measures are simply more likely to be introduced in less corrupt environments. If we encourage ACAs to promote transparency measures but we cannot show a direct causal effect on reduction of corruption, that does not mean that the ACAs are necessarily to blame. Strong claims about attribution and causality require strong understanding of the theoretical debate. Theory-based impact evaluations can in turn enrich the theoretical debate considerably.

Generally, the first question to ask is whether a change is observable. The second question is whether the observed change can be attributed to the intervention. Questions on causality, impact, and attribution are methodologically challenging. Answering them involves building a counterfactual scenario. When attribution is not possible to establish, a contribution analysis can be conducted instead. Rather than trying to assess the proportion of change that results from the intervention (attribution), a contribution analysis aims to demonstrate a more limited assertion, that is, whether or not the intervention evaluated is one of the causes of observed change. Contributions may be ranked but not quantified. The analysis takes a step-by-step approach, building a chain of logical arguments while taking into account alternative explanations and relevant external factors.

Designing a strong results-based management framework, theory of change, or logframe can help identify what the organisation can be held accountable for, given its resources and constraints. If resources are available, the ACA can be expected to produce the desired outputs. One can always hold the ACA accountable for producing the required outputs. At the outcome level, however, the results are no longer completely within the control of the ACA. If the logic model holds, the production of outputs should lead to the desired outcomes, but external factors may interfere. The outcome level is thus a grey area for accountability. ACAs are responsible, but only to a certain extent.

At the highest level, that of impact, which is often the level for the overall goal in logframes, the ACA is not directly accountable. It cannot normally be expected that achievement of the objective can be attributed specifically to actions of the ACA, only that the achievement of the ACA's outcomes contributes positively to the goal. In practice, this has serious implications. For example, should reduction in corruption levels be an impact/goal-level indicator, or should it be an outcome indicator? This depends on the type of ACA. All ACAs have reduction of corruption as an ultimate goal. However, for preventive-model ACAs, reducing corruption levels might not be something that they can influence, except through indirect processes of education and awareness. In this case, a reduction in levels of corruption should not be considered a possible outcome, only an overall goal to which the ACA will hopefully contribute over a long time period.

Two contrasting concepts are useful in this context: sphere of influence and sphere of concern. The sphere of concern for all ACAs is very wide, and extends to reducing corruption. However, the sphere of influence is not always as wide.<sup>20</sup>

The following example can illustrate the above points. Often, the ACA is responsible for implementing the national anti-corruption strategy. In such a case, indicators for that national strategy and for the ACA itself will most certainly overlap. But it is important to distinguish the success of the national anti-corruption strategy from the success of the ACA, as the strategy depends on a range of actors beyond the ACA. Thus, all elements of an anti-corruption strategy are normally within the ACA's *sphere of concern*, but some are probably not within its *sphere of responsibility*. All actors in the national anti-corruption strategy *contribute* towards its overall objective (typically reducing corruption levels), but none can be held directly *accountable* for failure or have success *attributed* to them at this highest objective level. To assign accountability or attribution we need to ensure that the objective is within the ACA's sphere of responsibility and that the ACA has the resources required to directly influence the results. This is normally done by focussing not on the overall objective, but on clearer, disaggregated sub-objectives (see indicator catalogue in annex 1).

### Box 9. Attribution analysis and contribution analysis

*Attribution analysis* aims to assess the proportion of observed change which can realistically be attributed to a specific intervention being evaluated. This involves building a counterfactual scenario, that is, one in which the intervention does not take place.

*Contribution analysis* aims to demonstrate whether or not the evaluated intervention is one of the causes of observed change. It may also rank the evaluated intervention among the various causes explaining the observed change.

*Source:* European Commission 2006a.

<sup>20</sup> These concepts are used in a specialised evaluation methodology called outcome mapping. Outcome mapping is based on the observation that "the complexity of the development process makes it extremely difficult to assess impact (especially for an external donor agency seeking attribution)." It tries to measure outcomes and changes in behaviour of direct beneficiaries rather than impact (Earl, Carden, and Smutylo 2001, 1–14).

The World Bank's concept of "most likely association" is a modification to attribution, making it easier for evaluators to deal with.<sup>21</sup> Proponents of such methodologies stress that they capture the complex messy realities better than more established methods that try to "force-fit" reality to the parameters of the methods.<sup>22</sup>

## 6.4 The principle of the counterfactual

Impact evaluation is a counterfactual method. To answer impact questions, one should eliminate other possible explanations for the impact observed in order to determine whether the ACA's interventions, and no other factors, explain the changes. Establishing counterfactual scenarios, in combination with baseline measurement, is essential to test for and eliminate alternative explanations.

Establishing counterfactual scenarios and baseline measures is time-consuming. Therefore, time should be invested from the outset in identifying only a few outcome-level indicators to be tracked. It is better to track a few outcome-level indicators, and if possible establish measures for a counterfactual scenario for them, than to measure all activities/outputs of the ACA.

The use of a counterfactual test, making use of a comparison group/entity that has not been targeted by the ACA, is at the heart of any impact evaluation. As illustrated by figure 4, if hard evidence on the ACA's impact is wanted, one should find a comparison group to represent a counterfactual scenario, that is, a group which has not been affected by the intervention. This makes it possible to estimate the net impact (change score) by comparing the treatment group with the non-treatment (control) group. This makes it possible to isolate the effect of the ACA by subtracting the comparison group value. A simple before-and-after analysis would lead to the conclusion that the ACA had been successful in producing I1. However, this is misleading, given that other external factors also had effects, causing the comparison group to reach C1. Therefore, the impact (added value) of the ACA is I1 minus C1.

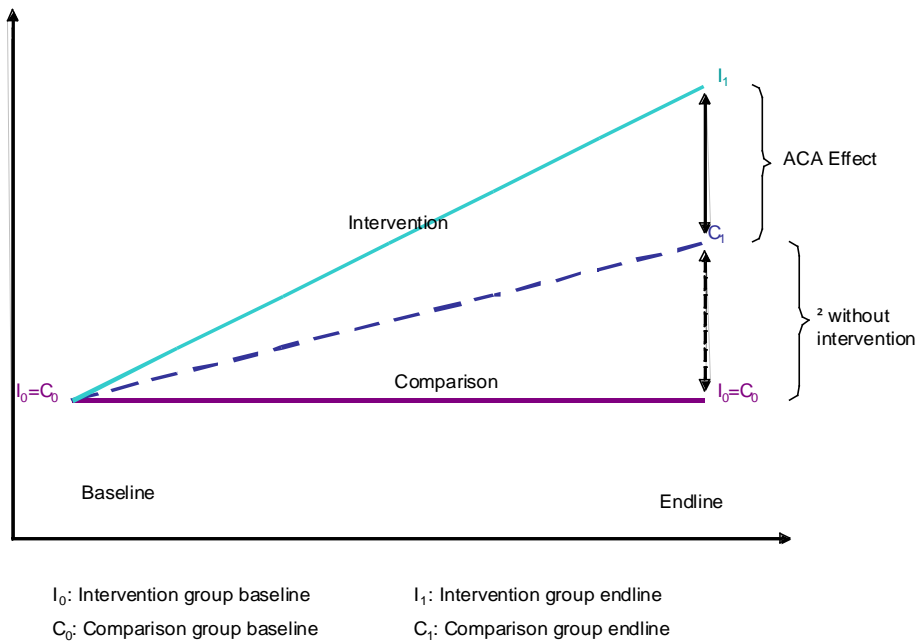
Because their target groups are most often nationwide, ACAs will often have difficulties identifying specific target groups and thus also in defining groups for the counterfactual scenario, that is, groups not being targeted by the intervention. Again, the likelihood of being able to create counterfactual scenarios increases with the disaggregation of the agency's goals and activities (into, for example, sector-specific activities). If a comparison group cannot be established to represent the counterfactual scenario, the status quo (absence of intervention) can be used instead.

Figure 4 shows a stylised impact evaluation approach. It shows how the aim is to "net out" the ACA effect by establishing what added value it produces compared to what would have happened without the intervention.

---

<sup>21</sup> In the words of the authors of the concept, it "allow[s] for a sound evaluative judgement based on the best evidence available while at the same time acknowledging that conditions are far from experimental and that data and knowledge gaps are widespread" (Foresti et al. 2007, 23).

<sup>22</sup> Readers interested in methodologies that aim to capture social change may wish to consult the literature on outcome mapping (see note 18) and on the "most significant change" technique. The latter is a more narrative-based approach than the traditional quantitative methods normally used to assign attribution (Davies and Dart 2005). This methodology and many other useful M&E resources are available on the Monitoring and Evaluation News website (<http://www.mande.co.uk>).

**Figure 4. Stylised impact evaluation approach**

Having an appropriate comparison group is, strictly speaking, a requirement for a rigorous impact evaluation. However, if this is not possible, a counterfactual scenario should still be constructed on the basis of hypothetical reasoning to estimate the likely effects of not making the intervention.

## 7. Indicators to track performance: Introducing the indicator catalogue

Indicator development requires a deep understanding of corruption and how individuals and institutions set out to fight it. Even a strong research methodology and thoughtful execution of the evaluation is unlikely to produce useful findings if the indicators are not finely attuned to reality. Therefore, we have devoted much effort to indicator development in this report. While the development of good indicators is the responsibility of the ACAs themselves in their planning cycle, evaluators and donors may play a constructive role by providing advice, quality assurance, and critique.

While it is true that measurement tools in the anti-corruption field are not as advanced as in other sectors such as health and education, a range of useful indicators do exist and can be combined to provide a clear picture of the performance of an ACA. The challenge for measuring progress in the fight against corruption is that off-the-shelf tool sets or sets of indicators are not appropriate, given the very context-specific nature of corruption. This catalogue of indicators in annex 1, therefore, presents a wide range of possible indicators, leaving open options for tailoring and flexibility.

### 7.1 Construction of indicators

To measure success one needs indicators. An indicator can be defined as a measure tracked systematically over time that indicates positive, negative, or no change with respect to progress towards a stated target. Managers or policy makers should examine the combined evidence from a group of indicators to evaluate whether the intervention is having positive effects. No outcome or impact should be measured by only one indicator. Indicators are normally derived from the impact, outcomes, and outputs defined in advance as desired results. It is therefore important to establish those clearly and to ensure that a clear logic or theory of change is established.

Indicators can be used for both monitoring and evaluation purposes, but they should be treated differently for each purpose. An evaluation can use the same indicators used for monitoring, but it requires a broader approach, in-depth analysis, use of other data sources, and a methodology adapted to the particular case. Nevertheless, monitoring data are often crucial for conducting successful evaluations, particularly those with the ambition of determining impact (Morra Imas and Rist 2009, 108). In the indicator catalogue, we present indicators that are applicable for monitoring purposes but which can also be used by evaluators.

Indicators should so far as possible be disaggregated to capture differences in, for example, types of corruption, corruption by sector (public, private, police, customs, etc.), gender, locality, methods of reporting corruption (e-mail, letter, personal, etc.). This is illustrated in the example below, from the Independent Commission Against Corruption (ICAC) in Hong Kong.

#### **Box 10. Indicators, constructs, and measures**

A *construct* is a general concept or idea used to define a result, e.g., good governance, environmental quality, etc. An *indicator* represents a key aspect of the construct; it is “indicative” of the construct although it may not represent it fully. A *measure* expresses an indicator’s value quantitatively or qualitatively. People frequently use indicator to mean much the same as measure; we use the terms separately to remind us that we should have *measurable indicators*.

We encourage use of mixed methods, both in the traditional sense of mixing quantitative and qualitative indicators, (e.g., quantitative statistics and qualitative case-studies) and also in an extended sense, mixing indicators which are perception-based or proxies with “harder” indicators which directly measure corruption (see below). There are several reasons for advocating mixed methods. Briefly, this approach provides a more comprehensive understanding of the phenomenon studied, allows triangulation of information sources, and reduces bias from a single data source.

Indicators can be categorised based on their purpose and on the type of measures they use.

The UNDP publication *A User’s Guide to Measuring Corruption* (2008, 9) outlines five types of corruption indicators:

- *Objective indicators:* These are indicators constructed from undisputed facts. Typical examples might include the existence of anti-corruption laws, the funding received by the anti-corruption agency, or national statistics.
- *Perception-based indicators:* Indicators may be based on the opinions and perceptions of corruption in a given country among citizens and experts.
- *Experience-based indicators:* These indicators measure citizens’ or firms’ actual experiences with corruption, such as whether they have been offered bribes or given bribes.
- *Proxy indicators:* On the premise that corruption is impossible to measure directly and empirically, proxy indicators assess corruption indirectly by aggregating “voices” and signals of corruption, or by instead measuring anti-corruption efforts, good governance, and public accountability.
- *Pro-poor and gender-sensitive indicators:* A pro-poor indicator requires a focus on people living in poverty, and a gender-sensitive indicator captures the different experiences and interests of women and men. Such indicators are useful for tracking the potentially different impacts that the mechanisms and processes of governance have on different social groups.

Whilst one can always argue about the terminology used for the different types of indicators, the essential point is that one should try to have a varied set of indicators, not solely relying on just one type such as perception-based indicators, for example. Unfortunately, purely empirical “objective” anti-corruption indicators based on hard facts are hard to find. As a result, the use of proxy indicators is very common in the social sciences. Many proxy indicators, if used correctly, can yield very good approximations to reality. However, one should always be aware of the nature of the indicator when interpreting it. The indicator of “the number of corruption cases brought to trial,” for example, should not be seen as a proxy for corruption levels in the country, as an increase in the number of cases brought to trial could indicate a higher incidence of corruption, an increased level of confidence in the court, or both; rather, it is a proxy for the efforts of the ACA and the judiciary. Similarly, when dealing with perception-based indicators, one should remember that they reflect people’s subjective opinions. A time lag will typically exist between any actual impact on corruption and a changed public attitude, and no direct causality can therefore be inferred from perception indicators (Galtung 2005). In sum, triangulation of different kind of indicators is needed to strengthen the validity and reliability of the findings.

The focus of the indicator catalogue is on indicator development, but we also briefly outline how data collection can take place, in the “sources of verification” column. There are several ways of measuring impact and keeping track of progress on selected indicators. Commonly used methods include public



opinion surveys, public sector diagnostics, private sector surveys, combined surveys, cross-country surveys, pro-poor and gender-sensitive surveys, and sectoral surveys (such as Public Expenditure Tracking Surveys).<sup>23</sup> National statistics can also be used if questions on corruption are incorporated into the data collection, enabling a comparison between statistics from different sources, comparable to the use of “mirror statistics” for trade data. Finally, the ACA can conduct its own data collection. This is referred to simply as “internal records” in the catalogue. An example is the “annual independent survey conducted by ICAC in Hong Kong which measures, among other things, the trust level between ICAC and the public, prosecution rate, as well as levels, types, location and causes of corruption” (UNODC 2002, 257).

## 7.2 Sample indicators used by three ACAs

The three ACAs below have been selected to provide a broad spectrum of organisations. The Serious Fraud Office (SFO) in the United Kingdom specialises in financial crime and has never received donor support. The Latvian Corruption Prevention and Combating Bureau (KNAB) used to be the recipient of donor-funded technical assistance, but it has gradually established organisational solidity and today lends expertise to other ACAs. Finally, some indicators for success of the Independent Commission Against Corruption in Hong Kong have been derived from its annual report and performance standards.

The indicators below are presented to illustrate how a few actual ACAs measure their performance. They constitute a selection rather than the full set of indicators used by these agencies, and the selection has been made without judging whether or not the indicators comply with the SMART or CREAM criteria (section 4.2.1). The indicator catalogue in annex 1, by contrast, has been created on the basis of a wider range of international experiences, including the work of the International Corruption Hunters Alliance gathering actual performance indicators from ACAs around the world. The indicator catalogue, moreover, is limited to indicators that comply with the principles for indicator development discussed above.

### 7.2.1 Serious Fraud Office, United Kingdom

Although the SFO in the United Kingdom has a narrower focus on financial crime than most ACAs in developing and transition countries, corruption is an important part of the organisation’s remit. With 307 permanent staff members, the SFO measures its performance through these indicators, among others:<sup>24</sup>

- Amount of budgetary allocation
- Total number of staff
- Number of active cases
- Price per UK inhabitant (used as a value-for-money measure)
- Number of trials
- Number of convictions
- Amount of money paid back to victims (financial compensation)

---

<sup>23</sup> A mapping of these different tools is provided by TI and UNDP (2007, 7–11).

<sup>24</sup> See the SFO’s website at <http://www.sfo.gov.uk/our-work/our-performance.aspx>, where the publication “Achievements in 2009–10” is available. See also the SFO’s annual report to Parliament 2009–2010, p. 23–25, at <http://www.sfo.gov.uk/media/112684/sfo%20annual%20report%202009-2010.pdf>.

- Amount of money recovered in corporate fines and channelled back into the court system (criminal financial recovery)
- Amount of money recovered from cases through civil means (civil financial recovery)
- Amount of money seized as part of financial investigations
- Number of answered requests for information from other countries for fraud and corruption cases (international cooperation)

### 7.2.2 Corruption Prevention and Combating Bureau, Latvia

KNAB in Latvia was established in 2002 and today has 142 staff members. Performance indicators for the bureau are available in its *Progress and Results in Preventing and Combating Corruption in Latvia* (KNAB 2009). Although this narrative report does not specifically formulate performance indicators as such, one can derive from it the following list:

- Number of legislative proposals in the area of anti-corruption
- Number of draft legal regulations presented
- Number of new laws successfully introduced
- Number of amendments to existing laws successfully adopted by parliament
- Amount of money spent illegally on political party financing discovered
- Amount of money spent illegally on political party financing reimbursed by parties
- Number of asset declarations analysed
- Amount of money earned by people due to breaking the laws on conflict of interest and additional employment and discovered by the ACA
- Amount of money recovered by the ACA
- Number of people convicted as a result of investigations by the ACA
- Number of criminal proceedings forwarded to the prosecutor's office
- Number of public servants trained on issues of conflict of interest, ethics, and internal anti-corruption measures
- Number of international requests for the ACA to provide its expertise abroad
- Number of hosted delegations from ACAs abroad

KNAB and the Lithuanian STT (Special Investigation Service), both ACAs, used the following indicators to measure their comparative performance in the report *Common Standards and Best Practices for Anti-corruption Agencies*, produced by the European Partners against Corruption Working Group (2008):

- Number of initiated investigations
- Number of cases sent to prosecution

### 7.2.3 Independent Commission Against Corruption, Hong Kong

Performance standards can also be made operational as indicators. The ICAC in Hong Kong, with approximately 1,200 employees, has the following performance standards, which can be translated

into indicators by counting either the number of times the standards have been met or the times they have not been met:<sup>25</sup>

- Respond to a report of corruption within 48 hours (indicator: ratio of number of corruption reports answered within 48 hours to number of corruption reports not answered within 48 hours)
- Respond to a report which does not involve corruption within two working days (indicator: ratio of number of non-corruption reports answered within two working days to number of non-corruption reports not answered within two working days)
- Respond to a request for corruption prevention advice within two working days (indicator: ratio of number of advice requests answered within two working days to number of advice requests not answered within two working days)
- Respond to a request for anti-corruption education services or information within two working days (indicator: ratio of number of service requests answered within two working days to number of service requests not answered within two working days)

Examples of other indicators that are not service-related but rather outcome-oriented were distilled from the annual report (ICAC 2009):

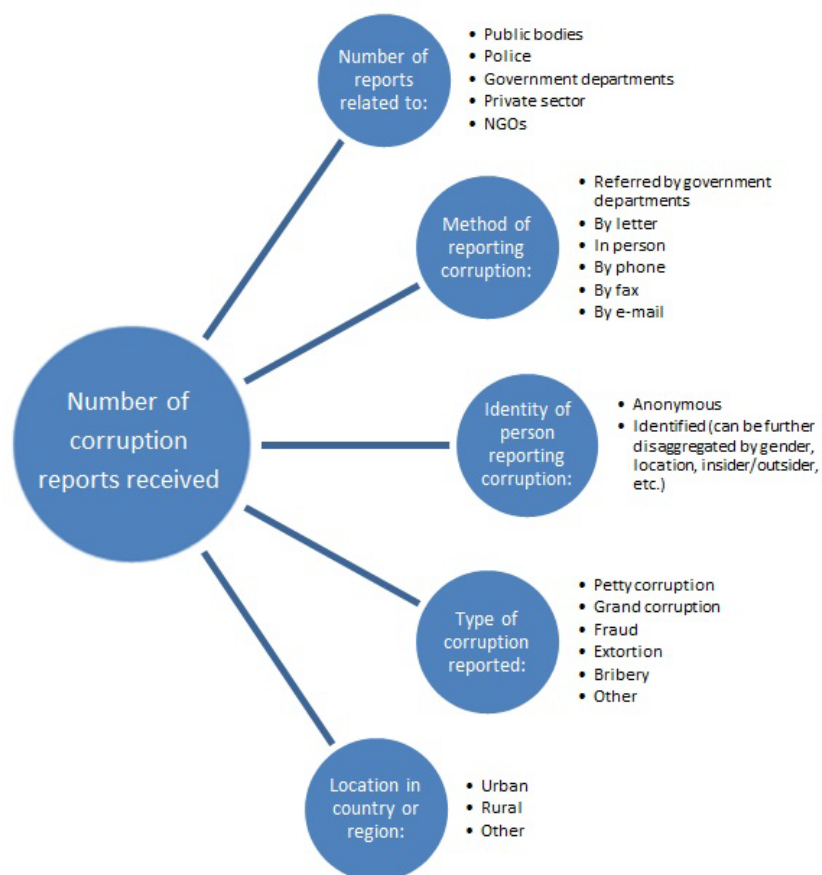
- Case-based conviction rate
- Number of visitors on Open Day
- Number of corruption reports received (disaggregated)
- Number of pursuable reports
- Number of investigations ongoing
- Number of investigations commenced, carried forward, and completed
- Total caseload
- Number of prosecutions (disaggregated by type of corruption)
- Percent of population having confidence in the ICAC
- Number of overseas visitors to the ICAC

Several additional indicators could be derived, such as indicators related to training and skills development or to corruption prevention and education (students taught, number of universities which have ethics as a component of studies, etc.).

The ICAC annual report also illustrates the benefits of disaggregating indicators to show important differences in the types of corruption, which enables a better understanding of the problem and thus means to combat corruption. Figure 5 shows an example of how one might disaggregate the indicator “number of corruption reports received.”

---

<sup>25</sup> See the ICAC website on performance standards, [http://www.icac.org.hk/en/about\\_icac/mp/index.html](http://www.icac.org.hk/en/about_icac/mp/index.html). Other performance indicators are available in ICAC (2011). [http://www.icac.org.hk/filemanager/en/Content\\_1017/pledge\\_eng.pdf](http://www.icac.org.hk/filemanager/en/Content_1017/pledge_eng.pdf)

**Figure 5. Disaggregation of sample indicator**

### 7.3 Guidelines for use of the indicator catalogue

It is important to remember that before selecting any of the indicators in the catalogue, one should consider whether they are relevant to the particular ACA in question and to the impacts/goals, outcomes, and outputs this organisation wants to produce. Indicators should not be chosen before the goals, outcomes, and outputs are clearly described.

Some ACAs may have specific mandates which require development of unique indicators. For example, ACAs with mandates related only to policy advice and coordination of the national integrity system would not be able to use indicators developed for law enforcement or preventive purposes. A set of indicators might also be chosen to focus on an ACA's specific area of weakness, regardless of whether this weakness is perceived or real. For example, if the ACA is highly productive in terms of administration, but not in enforcing the law, a key performance indicator might be the conviction ratio. If there is too much focus on petty corruption and major offenders are not being pursued, an indicator for better performance could be the percentage of cases where significant assets have been involved and/or seized (to create an incentive and show commitment to high-level cases).

The indicators in annex 1 have been categorized according to the basic functions which ACAs typically perform:

- *General management:* These are processes related to organisational performance and management practices, such as organisational strategy, standard operating procedures, internal oversight mechanisms, reporting procedures, and standardisation of work processes through documentation and guidelines (manuals, handbooks), etc.

- *Production, management, and sharing of knowledge:* The M&E system is a core component, but this function also includes analytical processes, management, and dissemination practices of all other departments.
- *Enforcement:* This function encompasses the three key enforcement processes of intelligence gathering and analysis, investigations, and prosecutions.
- *Legislation:* This includes analysis and improvement of national as well as international legislative documents related to the fight against corruption, benchmarking/compliance reviews based on international standards, and gap analysis of implementation of national legislation.
- *Prevention:* These processes are aimed not at enforcing the law once a criminal activity has been undertaken, but at preventing such activity. Civic education is a major component, as are various integrity measures such as codes of conduct, public integrity plans, asset declarations, etc.
- *Inter-agency cooperation:* ACAs often play a crucial role as focal points for various law enforcement, judicial, and other public sector institutions involved in the fight against corruption. A common element in promoting inter-agency cooperation is to improve case-processing procedures and practices.
- *International cooperation:* This embraces processes at the international level to stop transnational corruption, exchange information, and promote learning.
- *Civil society cooperation:* These are work processes involving civil society actors that do not fall within the above categories.
- *Business cooperation:* These are work processes involving business associations or individual corporations that do not fall within the above categories.

Most goals, outcomes, or outputs relate to such overall functions. Thus, if an ACA has a desired outcome relating to corruption prevention, one can narrow one's search in the matrix below by searching specifically for indicators relating to this function. Each indicator is also attached to a specific sub-function in order to clarify what aspect of the main function the indicator is meant to measure. Possible sources of verification are also listed for each indicator to show how data should be collected and analysed. Finally, comments and assumptions are presented for the indicators.

The indicator catalogue presents an array of options rather than a recommended mandatory list. It is important to select only those indicators that best reflect the functions and priorities of the ACA being evaluated. Few qualitative indicators are included. This is not because quantitative indicators are valued more highly than qualitative ones, but rather because qualitative indicators typically try to measure highly context-specific achievements less easily included in a general catalogue. These are often broken down into milestones.<sup>26</sup>

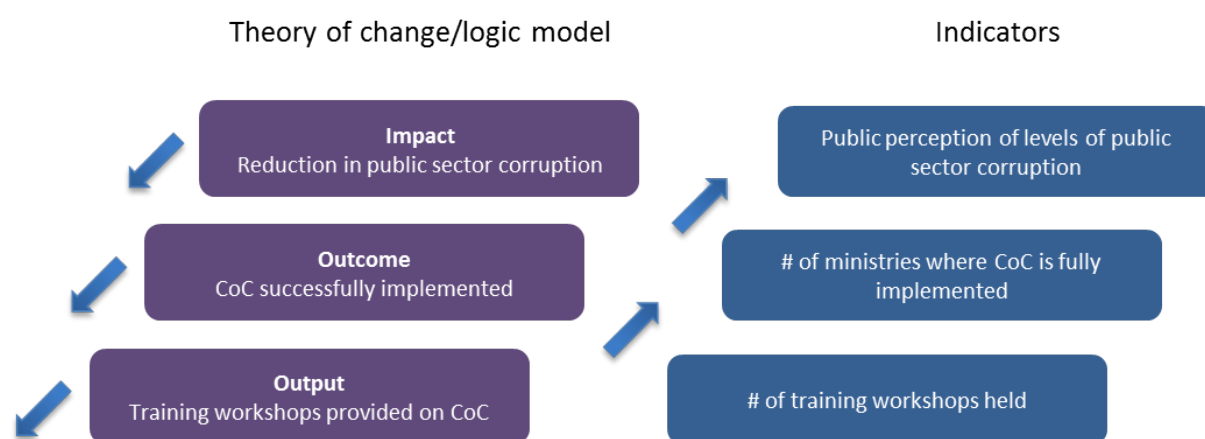
The catalogue in annex 1 is constructed on the logic of a normal logframe model, based on a theory of change. Hence, outputs lead to outcomes which lead to impacts. When constructing logframes and sets

---

<sup>26</sup> An example could be the adoption of conflict of interest legislation in parliament. Milestone 1: conflict of interest legislative input drafted for parliament group by ACA. Milestone 2: conflict of interest law adopted by parliament. Milestone 3: by-laws and government regulations produced on the basis of the law. Milestone 4: conflict of interest law implemented across public service. The DFID logframe is particularly useful when operationalising qualitative indicators (DFID 2009). <http://www.dfid.gov.uk/Documents/publications1/how-to-guid-rev-log-fmwk.pdf>

of indicators, it is important to establish the internal logic linking the output, outcome, and impact/goal levels. For the theory of change, this is typically done top-down: that is, we know what impact we want, and from there we decide which outcomes are necessary and which outputs are required to produce those outcomes. Indicators should be constructed the other way around, from the bottom up. One begins with the output indicators, then constructs the outcome indicators, and finally develops the impact-level indicators. This is of course a stylised view, since a strong internal logic is only developed through moving both upwards and downwards in the causal chain. But it is useful as a guiding principle. Figure 6 is based on the example of codes of conduct, but the same model could apply to any intervention.

**Figure 6. Linking logic model to indicators**



*Source: Adapted from slide presented at “Monitoring & Evaluation for Results,” training course organized by World Bank Institute Evaluation Group, Washington, DC, 2009.*

Figure 7 shows how the catalogue works, based on the principles above. As shown by the purple arrows, the desired impact determines the necessary outcomes and supporting outputs, and hence the functions which ACAs need to have. The blue arrows show that the indicators should be chosen/constructed with a bottom-up logic. The red arrows show the natural order of indicator development: first, one establishes the desired impact (at goal level) and functions (at outcome and output levels) one wants to measure. Then one establishes indicators, finds sources of verification, and explains the assumptions. For the outcome and output levels, there is an extra step, as the functions are first translated into desirable outcomes/outputs and then made operational in indicators. It is desirable for the purposes of this catalogue to categorise indicators under overall functions which ACAs normally have. However, this is not relevant for the goal/impact level. At the goal level, it is common to make use of available, existing, standard corruption measurement tools. Note that normally the goal level is considered outside the “sphere of influence” of the intervention, in other words, there is no direct attribution expected.

**Figure 7. How to use the indicator catalogue in annex 1**

IMPACT/GOAL LEVEL				
Desired impacts		Possible indicators	Sources of verification	Comments, assumptions
Impact 1		Indicator 1a Indicator 1b	Source 1a Source 1b	
OUTCOME LEVEL				
Functions	Desired outcomes	Possible indicators	Sources of verification	Comments, assumptions
Function 1	Outcome 1.1 Outcome 1.2	Indicator 1.1a Indicator 1.1b Indicator 1.2a Indicator 1.2b	Source 1.1a Source 1.1b Source 1.2a Source 1.2b	
Function 2	Outcome 2.1 Outcome 2.2	Indicator 2.1a Indicator 2.1b Indicator 2.2a Indicator 2.2b	Source 2.1a Source 2.1b Source 2.2a Source 2.2b	
OUTPUT LEVEL				
Functions	Desired outputs	Possible indicators	Sources of verification	Comments, assumptions
Function 1	Output 1.1.1 Output 1.1.2 Output 1.2.1 Output 1.2.2	Indicator 1.1.1a Indicator 1.1.1b Indicator 1.1.2a Indicator 1.1.2b Etc.	Source 1.1.1a Source 1.1.1b Source 1.1.2a Source 1.1.2b Etc.	
Function 2	Output 2.1.1 Output 2.1.2 Output 2.2.1 Output 2.2.2	Indicator 2.1.1a Indicator 2.1.1b Indicator 2.2.1a Indicator 2.2.2b Etc.	Source 2.1.1a Source 2.1.1b Source 2.2.1a Source 2.2.2b Etc.	

This logic should be followed when using the indicator catalogue to ensure that a coherent set of indicators are established. These should cover the impact, outcome, and output levels, and they should be appropriate for the intended functions of the ACA.

## 8. Learning from M&E

Over the past two decades the number of ACAs has grown considerably. Today, around 100 agencies are in operation all over the world. There are also many subnational entities, for which no numerical estimate is available. The character of these agencies varies greatly from one country to another. Some countries have endowed their agencies with investigative and even prosecution powers, whereas others have preferred a more preventive, educational, policy-related, or coordinative role. Some agencies are new while others have been in existence for some time. There have been many evaluations of ACAs, but what have we really learned?

The international community has been active in making ACAs a popular institutional response to corruption. The OECD (1996, 14) was the pioneer in suggesting the creation of this type of independent and specialized unit to member states as an integral part of their “ethics infrastructures.” Other initiatives then followed at the regional and international levels.<sup>27</sup>

However, even today, discussions on ACAs tend to follow a binary logic. ACAs are either seen as the answer, or not. There is great need for an informed debate focused not on *whether* ACAs are good or bad, but *why, to what extent, in which contexts*, etc. To inform an evidence-based debate, evaluations of ACAs should not only function as organisational audits, using a checklist or accountability approach to assess whether the organisations have all their prescribed powers and organisational structures. Rather, evaluations should try to elucidate the appropriate causal mechanisms for good anti-corruption interventions, clarify the actual impact of ACAs, and explore ways in which they can improve their work. In short, they should take a learning approach. Participation and the establishment of M&E units are both necessary (see sections 5.5 and 4.2). The following four components should be part of any M&E learning process:

- *Built-in feedback and dissemination mechanisms:* Evaluations should not be undertaken just to be shelved or seen only by a small group of people. The evaluation process should have built-in mechanisms for dissemination of the report and for feedback. When evaluations are done right, they contain important information for future policy and programme development. Here again, participation is vital to ensure a transparent process. Formal mechanisms for feedback could include scheduling review processes, peer reviews, seminars and workshops, etc. The ACA and donors should also encourage informal activities such as networking and internal communications that allow for the dissemination of ideas and information. As observed by the OECD, “In order to be effective, the feedback process requires staff and budget resources as well as support by senior management and the other actors involved” (2010a, 10). Therefore, feedback mechanisms should be planned and budgeted as part of the evaluation process; they should not be an afterthought. ACA and donor staff members who are responsible for integrating such feedback mechanisms into the evaluation process can consult the OECD (2001) publication *Evaluation Feedback for Effective Learning and Accountability*.
- *M&E as a learning tool:* Perhaps the biggest challenge is to promote a culture within the ACA that sees M&E as a useful learning tool rather than as an unwelcome duty. Participation is the first step towards making this happen. All ACA staff should be able to have their voices heard in the evaluation process and flag potential issues for attention.

---

<sup>27</sup> The Council of Europe Criminal Law Convention on Corruption, the Inter-American Convention against Corruption, the Southern African Development Community Protocol against Corruption, the African Union Convention on Preventing and Combating Corruption, and the United Nations Convention against Corruption (UNCAC) have all made reference to the need to establish independent authorities to combat corruption.



The next step is to ensure that the evaluation is relevant to the staff in their daily work, as noted below.

- *M&E should be useful:* Emphasising the idea that monitoring systems and evaluations should be useful might seem superfluous. However, if an evaluation does not have a defined purpose and designated users, then its usefulness is questionable. A good evaluation is designed, conducted, and reported with a sense of purpose and meets the needs of the intended users. Conclusions, recommendations, and lessons need to be clear, relevant, targeted, and actionable in order for the evaluation to promote learning (OECD 2010a, 29).
- *Learning from recommendations:* The ACA has a responsibility to ensure that the people responsible respond systematically to the recommendations of the evaluators. If ACA management personnel do not agree with the recommendations, they should voice their concerns during the feedback/review process. A formal management response and follow-up system (for example, through an action plan) should be developed to systematise implementation of the recommendations. The M&E unit, together with senior management, should ensure that all agreed follow-up actions are tracked to ensure accountability for their implementation. Later evaluations should review how management responded to the recommendations, whether they were implemented, and whether the desired effect was produced. If not, plausible explanations for the outcomes should be given.

In sum, learning is arguably the most overlooked aspect of M&E. There are strong pressures on ACAs for accountability, but these should not overshadow the necessity of learning what works and why.

## Annex 1. Indicator catalogue

IMPACT/GOAL LEVEL			
Desired impact	Possible indicators	Sources of verification	Comments, assumptions
Reduction in the crime of corruption and related offences	<p>Number of corruption-related crimes reported</p> <p>Number of convictions</p> <p>Number of disciplinary actions</p> <p>Levels of victimisation</p>	<ul style="list-style-type: none"> <li>▪ National statistics</li> <li>▪ Police reports</li> <li>▪ Court records court cases</li> <li>▪ Transparency International's Global Corruption Barometer</li> </ul>	<p>Can be further disaggregated by type of crime according to legal definition/type of corruption. When assessing court cases, need to clarify the role of ACAs through content analysis.</p> <p>Need to track this indicator over a long period. ACAs can be expected to contribute to higher number of cases, but success depends on many other actors.</p>
Improved transparency and less corruption in public service delivery	Percentage of community-level responses satisfied with performance of public service delivery in terms of transparency and corruption	<ul style="list-style-type: none"> <li>▪ Service delivery surveys</li> <li>▪ User satisfaction surveys</li> <li>▪ Expert opinions</li> </ul>	<p>The specific wording of the questionnaire will depend on the local setting and institution.</p> <p>Important to track same institutions over time for consistency. Need to focus ACA work on targeted institutions.</p>
Reduction in perceived corruption levels	Perception of corruption level by public opinion surveys	<ul style="list-style-type: none"> <li>▪ Public opinion surveys</li> <li>▪ Regional barometers (Africa, Asia, Latin America, Europe)</li> <li>▪ Transparency International's CPI, Bribe Payers Index, and Global Corruption Barometer</li> <li>▪ NGO reports</li> <li>▪ Media monitoring</li> </ul>	Measurement of general public perception of corruption levels
Cleaner business environment	Amount of money paid in bribes by	<ul style="list-style-type: none"> <li>▪ Global Competitiveness Index</li> </ul>	The assumption behind this indicator is that a business environment with less corruption

(greater freedom from corruption)	companies	<ul style="list-style-type: none"> <li>▪ PriceWaterhouseCoopers Global Economic Crime Survey</li> <li>▪ World Bank Business Environment and Enterprise Performance Survey (BEEPS)</li> <li>▪ World Bank &amp; IFC Enterprise Survey</li> <li>▪ US State Department Investment Climate Survey</li> </ul>	facilitates economic growth.
Greater interpersonal trust, leading to reduced corruption	Levels of interpersonal trust	<ul style="list-style-type: none"> <li>▪ Quality of Government Institute Survey Dataset</li> </ul>	Various measures of interpersonal trust exist which need to be specified. Could also be combined with trust in government institutions, at outcome level.
Public awareness of the negative effects of corruption	Awareness level among population of negative consequences of corruption, as shown in public opinion surveys	<ul style="list-style-type: none"> <li>▪ Public awareness surveys</li> </ul>	This seeks to measure specifically whether the population is aware of the negative consequences of corruption.

OUTCOME LEVEL				
Function	Desired outcomes	Possible indicators	Sources of verification	Comments, assumptions
General management	Performance is viewed as good by donors	Donor reports recognising positive contribution by the ACA (yes/no)	<ul style="list-style-type: none"> <li>▪ World Bank Governance and Anti-Corruption (GAC) Diagnostic Surveys</li> <li>▪ Group of States Against Corruption (GRECO) reports</li> <li>▪ External evaluations</li> </ul>	At the overall outcome level, ACA performance can be assessed by expert evaluators, typically provided by donor agencies.
	Performance is viewed as good by population	Percentage of population who find the performance of the ACA satisfactory	<ul style="list-style-type: none"> <li>▪ Surveys</li> </ul>	Public satisfaction with the ACA is important, but it is important to note that it takes many years to build up knowledge about an ACA's work in the general population.
	Public has confidence in the ACA	Level of public trust in the ACA's commitment to fight against corruption	<ul style="list-style-type: none"> <li>▪ Surveys</li> </ul>	This is a subset of the previous indicator. Since the level of trust between the public and the ACA is critical for the success of anti-corruption efforts, public trust levels should be monitored. Could be disaggregated to measure specific integrity measures, such as trust in the commissioner.
	The media perceive the ACA to be competent and achieving results	Number of positive and negative press stories, whether case-related or general information, on the ACA	<ul style="list-style-type: none"> <li>▪ Media monitoring, content analysis of press articles</li> </ul>	In countries where the media are relatively independent, an indicator of ACA performance can be its treatment in the media. There can be disaggregation of press stories using content analysis into, e.g., high-profile versus low-profile cases. Comparing successful and unsuccessful cases would give much richer information.
	ACA organisational structure, systems, and processes are	Number of revealed shortcomings in expert	<ul style="list-style-type: none"> <li>▪ Expert review</li> </ul>	Assessments could review whether job functions are clearly described and

	functioning well	assessment reports Number of shortcomings identified as “critical” and requiring immediate action	assessments	linked to an organogram, and review job descriptions, the hiring and promotion process, etc.
	ACA has competent staff	Staff competence levels in anti-corruption-related areas	<ul style="list-style-type: none"> <li>Comparison of individual competence levels using tests before and after</li> </ul>	This depends on the kind of training activities provided.
	ACA provides useful training for its staff	Staff satisfaction levels with training delivery, training curricula and materials, and training modules	<ul style="list-style-type: none"> <li>Training feedback forms</li> </ul>	The precise formulation of this indicator depends on the training arrangements of the ACA. If in-house trainers are used, then training satisfaction is a good indicator.
Knowledge production and management	ACA has good diagnostic and analytical capacity	Number of studies performed or commissioned by the ACA	<ul style="list-style-type: none"> <li>Internal records</li> </ul>	The number of studies is a crude indicator of the quality of the diagnostic capacity of the ACA. More sophisticated assessment would be needed for evaluation purposes.
	ACA has an operational, results-based M&E system and capacity to run it	M&E system established and operational (yes/no)	<ul style="list-style-type: none"> <li>Internal records/external expert assessment</li> </ul>	The use of “operational” in the indicator should at a minimum be defined to mean indicators established, data routinely collected, and a person responsible for collection and storing data in a systematic fashion.
	ACA has a functional and secure system for knowledge management and sharing	System for knowledge management and sharing established and operational, with adequate provisions for data protection (yes/no)	<ul style="list-style-type: none"> <li>Internal records/external expert assessment</li> </ul>	This includes assessment of how information is shared within the organisation; how staff access non-sensitive information; whether regular meetings are conducted; whether and how managers share information internally and with staff, through inter-office circulars or other means; whether there are database systems and clear identification of responsibilities to feed data into such systems, who can access them, etc.

Enforcement	Higher quantity and quality of prosecution of corruption cases	Number of corruption cases that are reported by the ACA and forwarded to the prosecutor's office or special court	<ul style="list-style-type: none"> <li>Official police, court, and prosecution statistics</li> </ul>	This indicator can be disaggregated if more details are desired on each stage of the process or type of cases. For example, it would be useful to know the type of legal offences forwarded to understand whether the ACA is successful in prosecuting high-level corruption.	
	Better integration of enforcement and prevention within ACA work processes	Number of managerial initiatives implemented and/or prevention activities initiated on the basis of data from enforcement division	<ul style="list-style-type: none"> <li>Internal records: review of annual reports, strategic plans, departmental work plans, etc.</li> </ul>	This indicator focuses on whether information from enforcement work (intelligence, investigation, prosecution) is used for initiating specific and/or generic institutional and procedural reforms, prioritisation, and strategic decisions, or to inform prevention and education activities.	
	Investigation capacity is better quantitatively and qualitatively	Investigations-to-prosecutions ratio Investigations-to-convictions and prosecutions-to-convictions ratios		<ul style="list-style-type: none"> <li>Internal records</li> <li>Court records</li> </ul>	This indicator measures not only the effectiveness of the investigators but also the effectiveness of the ACA's case-selection system
		Number of active cases under investigation by the ACA		<ul style="list-style-type: none"> <li>Internal records</li> </ul>	This indicator measures the workload of the ACA in terms of investigations.
		Number of cases dismissed or people acquitted due to procedural flaws or incompetent investigation by the ACA		<ul style="list-style-type: none"> <li>Official police, court, and prosecution statistics</li> </ul>	The number of serious procedural mistakes or poor investigatory practices by the ACA should be minimised. This indicator may also reflect the interaction between prosecution/courts and the ACA. If the relationship is not cooperative the indicator is not good.
		Number of cases dismissed or people acquitted due to lack of evidence		<ul style="list-style-type: none"> <li>Official police, court, and prosecution statistics</li> </ul>	This indicator does not necessarily reflect poor performance by the ACA (evidence cannot always be established, for example, if people are innocent) but it provides useful information.

	This level of convictions is high, including convictions of non-petty crimes	Number of convictions or percentage of convictions as proportion of total investigations	<ul style="list-style-type: none"> <li>▪ Court records</li> </ul>	This indicator really only reflects ACA performance if the agency has prosecutorial powers and the courts work well.
		Percentage of convictions where significant assets have been involved and/or seized	<ul style="list-style-type: none"> <li>▪ Court records</li> </ul>	This indicator focuses on high-level or grand corruption.
	Higher degree of asset seizure	Value of assets seized as part of ACA investigations	<ul style="list-style-type: none"> <li>▪ Internal records</li> </ul>	Rather than focusing on the number of seizures, the total amount of money provides a measure for the overall impact (100 small seizures may not be as significant as one large one).
	Better asset recovery system	<p>Number of shortcomings identified in expert assessment of asset recovery systems</p> <p>Number of shortcomings identified as “critical,” requiring immediate action</p>	<ul style="list-style-type: none"> <li>▪ Internal documentation/ external expert assessment</li> </ul>	Assessments would benefit from covering typical weaknesses of asset-recovery systems: how assets are evaluated; how they are managed; what possibilities are provided by the law to liquefy assets due to likely depreciation; whether there is a clear distinction between seizure and confiscation/forfeiture; whether assets are double- or triple-counted at various stages of the procedure (e.g., seizure, confiscation, disposal); what happens with the proceeds after disposal; whether the public benefits in any visible way from the recovered assets.
	ACA contributes to financial recovery of proceeds of crime	Amount of money recovered in fines and compensation money channelled back via the court system	<ul style="list-style-type: none"> <li>▪ Internal records</li> <li>▪ Court records</li> </ul>	This measures the financial benefits either for the state or for individuals through proceeds of crime recovered.
Legislation	New domestic laws have been introduced	Number of new anti-corruption legislative proposals successfully introduced into law	<ul style="list-style-type: none"> <li>▪ Official parliamentary records/gazette</li> <li>▪ Internal records</li> </ul>	This indicator can also be changed into a more qualitative indicator, outlining the specific areas where new laws are needed.

	Domestic laws have been amended	Number of amendments to existing laws successfully adopted by parliament	<ul style="list-style-type: none"> <li>Official parliamentary records/gazette</li> </ul>	Again, rather than the number of amendments, one could list the specific amendments that the ACA wants to work on as targets.
Policy	ACA has contributed policy inputs to the national agenda	Number of policy inputs produced by the ACA	<ul style="list-style-type: none"> <li>Internal records</li> </ul>	Policy inputs can be defined as policy proposals, advocacy letters, media inputs, etc.
	ACA has had positive policy influence	Number of government policy documents/laws per year that are measurably influenced by inputs from the ACA	<ul style="list-style-type: none"> <li>Internal records</li> <li>Official records</li> <li>Newspaper articles</li> </ul>	“Measurably” implies that a contribution to the policy change/enactment can be attributed to the ACA, via objective measures, such as public records, news reports, or reflection of the policy change in the ACA’s advocacy or strategic plan.
Prevention	ACA promotes a strong public service ethos	Percentage of public servants who are aware of ethical dilemmas and guidelines regarding corruption	<ul style="list-style-type: none"> <li>Survey</li> <li>Training feedback forms</li> </ul>	The assumption is that the ACA or other institutions provide courses in ethics and anti-corruption.
		Percentage/number of full investigations of significant breaches of procedures or ethics of senior civil servants being pursued fully and fairly and leading to transparent outcomes, and if necessary, to enforcement/punishment in institutions targeted by the ACA	<ul style="list-style-type: none"> <li>Ministry records</li> </ul>	Unless this indicator is already measured by the civil service, it requires much work from the ACA to measure. If public sector corruption is a key issue this work would be worth undertaking, but ultimately it is a matter of prioritisation of resources.
	ACA promotes trust in public service	Percentage of public service users who have trust in the institutions	<ul style="list-style-type: none"> <li>Survey</li> <li>Customer satisfaction survey</li> </ul>	This indicator is based on perception/experience, and one should therefore not expect quick changes. The indicator could also be specifically targeted to only those public service institutions which the ACA works with



	Reduced time/real costs of obtaining licensing or utilities	Time/real cost to service users of getting a license from a selected licensing agency or connection to utilities through a government utilities company	<ul style="list-style-type: none"> <li>▪ Survey</li> </ul>	Real cost includes the official price plus average kickback required.
	ACA combats illegal political party financing	Amount of money spent illegally on political party financing disclosed by ACA	<ul style="list-style-type: none"> <li>▪ Internal records</li> </ul>	If political party financing is an important issue to tackle, the amount of money disclosed is important to track.
		Amount of money spent illegally on political party financing reimbursed by parties	<ul style="list-style-type: none"> <li>▪ Internal records</li> <li>▪ Treasury records</li> </ul>	The amount of money reimbursed by political parties indicates both enforcement and preventive effects of the ACA's work.
	Reduced conflict of interest issues in public administration	Amount of money earned by people due to breaking the laws on conflict of interest and additional employment disclosed by the ACA	<ul style="list-style-type: none"> <li>▪ Internal records</li> </ul>	An alternative indicator could be the number of conflict of interest offences.
	Asset declaration works as corruption deterrent	Compliance rate for civil servants required to file asset disclosures	<ul style="list-style-type: none"> <li>▪ Internal records</li> </ul>	This indicator could also include measures on whether disclosure forms are properly filled out. The indicator assumes that the ACA is in charge or oversees asset declarations. An indicator of the quality of the ACA's work could be done via random checks on the accuracy of asset declaration forms.
	Corruption risk audits and/or public integrity plans promote lower levels of corruption	<p>Number/percentage of public institutions that have implemented the majority of recommended measures proposed by risk audits/integrity plans within suggested time frame</p> <p>Public/civil servants' perception of effectiveness of the risk audits/integrity plans</p>	<ul style="list-style-type: none"> <li>▪ Surveys</li> </ul>	This indicator is very context-specific and should therefore be tailored to the individual case.

Inter-agency cooperation	ACA has capacity to promote cooperation between institutions involved in the fight against corruption	Distribution of competences among anti-corruption institutions is clear, agreed upon, and conducive to fighting corruption (yes/no)	<ul style="list-style-type: none"> <li>▪ External assessments</li> <li>▪ Existence of manual/guidelines/inter-agency brief which specifies competences and cooperation procedures</li> </ul>	This qualitative indicator can be further operationalised with milestones representing the activities required to achieve this ultimate target.
	ACA actively cooperates with other anti-corruption institutions	Number of joint initiatives/operations/investigations developed and conducted	<ul style="list-style-type: none"> <li>▪ Internal records</li> </ul>	This of course depends on whether the ACA has investigatory powers.
	Case processing is enhanced	Average number of days from when a case is filed to when it reaches the courts Guaranteed maximum number of days which the case-processing process is allowed to take	<ul style="list-style-type: none"> <li>▪ Internal records</li> <li>▪ Court records</li> <li>▪ Inter-agency cooperation agreements</li> </ul>	Many variations of these indicators can be tailored to specific settings. For instance, if certain bottlenecks have been identified, indicators could focus on these.
International cooperation	Useful exchange of information takes place at the international level	Number of requests for information related to fraud and corruption from other countries answered by the ACA	<ul style="list-style-type: none"> <li>▪ Internal records</li> </ul>	It is an assumption that the ACA receives requests for information from foreign institutions, but this is an appropriate assumption at the outcome level.
	Joint operations are developed and conducted at the international level	Number of joint initiatives/operations developed and conducted with foreign anti-corruption institutions	<ul style="list-style-type: none"> <li>▪ Internal records</li> </ul>	This of course depends on whether the ACA has investigatory powers.
Civil society cooperation	Civil society has the capacity to hold the government accountable for corruption	Number of NGOs with specialized capacity to analyse, monitor, and publicize government corruption	<ul style="list-style-type: none"> <li>▪ Self-assessment by NGOs</li> </ul>	The assumption is that the ACA has provided these NGOs with training.

	The media are independent and free	Number of journalists murdered, expelled, and/or harassed because of reporting on corruption	<ul style="list-style-type: none"> <li>▪ Internal records</li> <li>▪ Reporters Without Borders data</li> </ul>	There are many international free media indexes which have indicators and scores for a range of countries. If working with the media is a priority for the ACA, it can also undertake its own data collection.
		State monopoly on radio and television (yes/no)	<ul style="list-style-type: none"> <li>▪ Internal records</li> <li>▪ Reporters Without Borders data</li> <li>▪ Freedom House data</li> </ul>	
		Number of censored media programmes/articles	<ul style="list-style-type: none"> <li>▪ Internal records</li> <li>▪ Reporters Without Borders data</li> </ul>	
Business cooperation	ACA contributes to a more corruption-free business environment	Number of corruption reports from business which are investigated by the ACA	<ul style="list-style-type: none"> <li>▪ Internal records</li> </ul>	It is important to only measure issues which the ACA can influence at the outcome level.
		Number of referrals from the ACA to other justice agencies on business corruption	<ul style="list-style-type: none"> <li>▪ Internal records</li> </ul>	Ideally one should also measure the end results of such referrals.
		The business community's perception of the ACA's contribution to curbing corruption	<ul style="list-style-type: none"> <li>▪ Surveys</li> </ul>	This can be done through existing surveys or locally.

OUTPUT LEVEL				
Function	Desired outputs	Possible indicators	Sources of verification	Comments, assumptions
General management	ACA is independent with secure budget	Amount of funding received by the ACA	Internal records	Any ACA needs sufficient funding to perform its tasks. While this is not always within the ACA's "sphere of influence," attracting funding is an important task and thus an appropriate indicator.
	ACA provides its staff with the necessary training	Number of training needs assessments (TNA) undertaken	TNA reports	At a later stage, a standardised TNA report can measure improvement in skill levels.
		Number of trainings undertaken and completed by ACA staff	Internal records	This is a rather crude measure of training as it does not measure the quality of the training. It should therefore be combined with the indicators below.
		Number/percentage of ACA staff who experience improved skill levels	Facilitated self-assessment through annual appraisal	Self-assessment forms can be used, but facilitated self-assessment can help reduce bias.
		Number/percentage of staff who complete specialist certificate, diploma, or master's-level training in anti-corruption	Internal records	This indicator could be combined with staff development plans of the agency for more specific sub-indicators.
		Number of ACA staff who have conducted study visits to review anti-corruption best practices	Internal records	If specific plans for study visits are already in place, this can be transformed into a qualitative indicator with specific study visits as milestones.
		ACA has adequate staffing levels (appropriate staffing number to be agreed between ACA, donors, and government, based on functions, tasks, and expectations for the ACA)	Number of staff	Internal records
	The agency is fully staffed (yes/no)		Internal records	If a staffing target has been set, this can be used as a qualitative indicator. It can be broken down into achievable milestones.
	Recruitment rates		Internal records	Another indicator for staffing is recruitment rates.
	Retention rates		Internal records	As time passes, retention rates may be more important than recruitment rates.

	IT equipment and software is adequate for the ACA to perform its work	ACA's needs in terms of IT equipment and software applications have been reviewed and a plan developed and implemented (yes/no)	Internal records	This qualitative indicator can be broken down into milestones.
	ACA is able to communicate its successes and interact with society	Number of visits/downloads/length of stay/interactions on the agency's website	Internal records	This is a good indicator of the usefulness of the website in countries where Internet access is widespread.
		Number of information meetings and events held by the ACA	Internal records	This can include meetings with journalists, communities, youth organisations, etc.
	ACA is service-minded and responsive	Percentage of responses to a corruption report within predefined number of hours	Internal records	Can be used as an internal indicator and motivator for service to citizens.
		Percentage of responses to requests for corruption prevention advice within predefined number of working days	Internal records	Can be used as an internal indicator and motivator for service to citizens.
Legislation	ACA produces legislative inputs on anti-corruption	Number of legislative inputs for drafting laws prepared by ACA	Internal records	It cannot be expected that the ACA will draft laws for parliament, but a performance indicator is whether the ACA has prepared constructive inputs for this process.
	Compliance reviews of anti-corruption legislation are conducted by the ACA	Number of compliance reviews undertaken	Internal records	This could be disaggregated by sectors or international conventions.
Enforcement	ACA has systems in place to promote receipt of corruption reports	Number of corruption reports received	Internal records	Can be disaggregated (as in the ICAC example above). One useful distinction might be between reports on petty corruption versus grand corruption.
	ACA is able to investigate corruption cases successfully	Number of investigation cases initiated	Internal records	An indicator of investigation output, not outcome.
Prevention	ACA promotes strong public service ethos	Number of public servants trained on issues of conflict of interest, ethics, and internal anti-corruption measures	Internal records	A purely quantitative measure of number of trained civil servants. If different kinds of trainings are provided, the indicator can be disaggregated according to types.

	ACA contributes to development and implementation of codes of conduct	Number of public service institutions that have developed and implemented codes of conduct	Internal records	The assumption is that the ACA has provided advice on the codes of conduct.
	ACA contributes to development and implementation of public integrity plans	Number of public service institutions that have developed and implemented public integrity plans	Internal records	The assumption is that the ACA has advised on the public integrity plans.
	Assets of public officials are declared and analysed	Number of asset declarations analysed	Internal records	This is a crude indicator and it is not necessarily a measure of good performance to have analysed a large number of indicators. The most important indicators regarding assets are the outcome indicators (see above).
	ACA contributes to improvement of audits	Percentage of government budget audited to required standards in the last financial year	Statistics from auditor general's office	The ACA is typically not directly responsible for audits, but if this is a priority area and the ACA works with other institutions to improve this area, the indicator is valid. However, no causal effects can be attributed to the ACA.
		Percentage/number of audits that reveal inconsistencies	Statistics from auditor general's office	The assumption is that the ACA contributes to better audits through capacity building.
	ACA contributes to better procurement	Percentage of government contracts and procurements reviewed in the last financial year	Government statistics	Procurement is rarely a responsibility of the ACA, but if a concentrated effort is made to improve work on procurements, this indicator is valid. However, no causal effects can be attributed to the ACA.
		Percentage/number of government inspections of contracts and procurements that reveal inconsistencies	Government statistics	As above, only this indicator focuses on the result of the reviews.
	ACA raises public awareness of the negative effects of corruption	National awareness campaign planned and launched (yes/no)	Internal records	A qualitative indicator which can be broken into milestones.
		Number of workshops and seminars organised for journalists, NGOs, and public administration, covering the	Internal records	This indicator can be further disaggregated according to, for example, target audience.

		ACA's mandate		
Inter-agency cooperation	ACA has capacity to promote cooperation between institutions involved in the fight against corruption	Number of conferences successfully completed	Internal records	This can be operationalised into a qualitative indicator by specifying when and how conferences should happen.
	ACA conducts joint, intergovernmental workshops	Number of joint intergovernmental workshops completed	Internal records Minutes from meetings	Holding regular joint workshops between institutions involved in the fight against corruption is assumed to improve coordination and cooperation.
Knowledge production and management	ACA has capacity to run M&E system	Presence of M&E system (yes/no)	Internal records External expert assessment	A qualitative indicator which can be broken down into sub-indicators.
	ACA collects data is collected in a systematic and regular fashion	Presence of person responsible for collecting and storing monitoring information (yes/no)	Internal records	This person would normally be an M&E staff member.
		Data routinely compiled for all chosen indicators (yes/no)	Internal records	A definition of "routinely" needs to be provided by the ACA depending on context and staff resources. Quarterly or half-yearly would be typical.
	ACA conducts research and analytical studies, such as risk assessments, on corruption	Number of research reports produced	Internal records	This indicator does not measure the quality of the research reports. It is assumed that reports of poor quality are not accepted.
		Number of corruption risk assessment reports produced	Internal records	The indicator focuses on risk assessment but other specialised areas of analytical work could be chosen.
International cooperation	ACA contributes to exchange of expertise internationally	Number of international requests for the ACA to provide its expertise abroad	Internal records	The assumption is that the number of requests from similar institutions to learn from the practices and personnel of the ACA corresponds to its professional reputation.
		Number of delegations hosted from ACAs abroad	Internal records	If delegations from similar institutions travel to the ACA's premises to learn, one can again assume that it indicates a high professional reputation.
		Number of overseas visitors to the	Internal records	This indicator measures individuals rather than

		ACA		institutions, including individual professionals, academics, donor staff, etc.
Civil society cooperation	ACA provides information/education to civil society	Number of meetings/workshops with journalists, youth leaders, NGO representatives, etc.	Internal records	Meetings are broadly defined. The indicator measures the outreach activity level of the ACA.
	Civil society organisations (NGOs, schools, unions, etc.) request and adopt materials from ACA	Number of civil society organisations requesting and adopting codes of conducts, ethics guidelines, etc.	Internal records	This depends on the initiatives and materials made available by the ACA.
Business cooperation	ACA provides information/education to businesses on how to avoid corruption	Number of meetings/workshops with businesses, business associations, etc.	Internal records	Meetings are broadly defined. The indicator measures the outreach activity level of the ACA
	Corporations request and adopt codes of conducts, ethics guides, and other materials recommended by the ACA	Number of corporations requesting and adopting codes of conducts, ethics guides, etc.	Internal records	This depends on the initiatives and materials made available by the ACA.



## References

- Arndt, Christine, and Charles Oman. 2006. *Uses and Abuses of Governance Indicators*. Paris: OECD Development Centre.
- Bamberger, Michael. 2006. *Conducting Quality Impact Evaluations under Budget, Time and Data Constraints*. Washington, DC: World Bank Independent Evaluation Group.
- Bamberger, Michael, Vijaendra Rao, and Michael Woolcock. 2010. "Using Mixed Methods in Monitoring and Evaluation: Experiences from International Development." In *Handbook of Mixed Methods Research*, 2nd ed., edited by Abbas Tashakkori and Charles B. Teddlie, 613–42. Thousand Oaks, CA: Sage.
- Bamberger, Michael, Jim Rugh, and Linda Mabry. 2006. *Real World Evaluation: Working under Budget, Time, Data and Political Constraints*. Thousand Oaks, CA: Sage.
- Bamberger, Michael, and Howard White. 2007. "Using Strong Evaluation Designs in Developing Countries: Experiences and Challenges." *Journal of Multidisciplinary Evaluation* 4, no. 8 (October): 58–73.
- Banerjee, Abhijit, Rema Hanna, and Sendhil Mullainathan. 2009. "Corruption." [http://www.hks.harvard.edu/fs/rhanna/Corruption\\_Chapter\\_Full8.pdf](http://www.hks.harvard.edu/fs/rhanna/Corruption_Chapter_Full8.pdf).
- Bardhan, Pranab. 1997. "Corruption and Development: A Review of Issues." *Journal of Economic Literature* 35 (September): 1320–46.
- Bjorkman, Martina, and Jacob Svensson. 2009. "Power to the People: Evidence from a Randomized Field Experiment of a Community-Based Monitoring Project in Uganda." *Quarterly Journal of Economics* 124, no. 2: 735–69.
- Boardman, Anthony E., David H. Greenberg, Aidan R. Vining, and David L. Weimer. 2006. *Cost-Benefit Analysis*. 3rd ed. Upper Saddle River, NJ: Pearson International.
- Bryman, Alan. 2004. *Social Research Methods*. 2nd ed. Oxford, UK: Oxford University Press.
- Byrne, Elaine, Anne-Katrin Arnold, and Fumiko Nagano. 2010. *Building Public Support for Anti-Corruption Efforts: Why Anti-Corruption Agencies Need to Communicate and How*. Washington, DC: World Bank.
- Center for Global Development. 2006. *When Will We Ever Learn? Improving Lives through Impact Evaluation*. Washington, DC: Center for Global Development.
- Davies, Rick, and Jess Dart. 2005. *The 'Most Significant Change' (MSC) Technique: A Guide to Its Use*. Cambridge, UK: Monitoring and Evaluation News. <http://www.mande.co.uk/docs/MSCGuide.htm>.
- Denzin, Norman K. 1970. *The Research Act in Sociology*. London: Butterworths.
- De Sousa, Luís. 2009a. *Anti-Corruption Agencies: Between Empowerment and Irrelevance*. Florence: European University Institute.
- . 2009b. *Does Performance Matter to Institutional Survival? The Method and Politics of Performance Measurement for Anti-Corruption Agencies*. Florence: European University Institute.
- De Sousa, Luís, Barry Hindness, and Peter Larmour, eds. 2009. *Governments, NGOs and Anti-Corruption: The New Integrity Warriors*. London: Routledge.
- De Speville, Bertrand. 2000. "Why Do Anti-Corruption Agencies Fail?" In "Implementation Tools: Inputs for a United Nations Experts Group Meeting" (unpublished). UN Global Programme Against Corruption (GPAC) and United Nations Centre for International Crime Prevention (UNCICP), Vienna.

- DFID (UK Department for International Development). 2009. *Guidance on Using the Revised Logical Framework*. How to Note: DFID Practice Paper. London: DFID. <http://www.dfid.gov.uk/Documents/publications1/how-to-guid-rev-log-fmwk.pdf>.
- Doig, Alan. 1995. "Good Government and Sustainable Anti-Corruption Strategies: A Role for Independent Anti-Corruption Agencies." *Public Administration and Development* 15, no. 2: 151–65.
- . 2000. "Getting the Boring Bits Right First: Capacity Building for Anti-Corruption Agencies (ACA)." In "Implementation Tools: Inputs for a United Nations Experts Group Meeting" (unpublished). UN Global Programme Against Corruption (GPAC) and United Nations Centre for International Crime Prevention (UNCICP), Vienna.
- . 2009. "Markets, Management, Managed Work and Measurement: Setting the Context for Effective Anti-Corruption Commissions." In *Governments, NGOs and Anti-Corruption: The New Integrity Warriors*, edited by Luís De Sousa, Barry Hindess, and Peter Larmour. London: Routledge.
- Doig, Alan, David Watt, and Robert Williams. 2005. *Measuring 'Success' in Five African Anti-Corruption Commissions: The Cases of Ghana, Malawi, Tanzania, Uganda & Zambia*. U4 Report. Bergen, Norway: U4 Anti-Corruption Centre.
- . 2007. "Why Developing Country Anti-Corruption Agencies Fail to Develop: Understanding the Dilemmas of Organisational Development, Performance Expectation and Donor and Government Cycles in African Anti-Corruption Agencies." *Public Administration and Development* 27, no. 3: 251–59.
- Duncan, Nick. 2006. "The Non-Perception Based Measurement of Corruption: A Review of Issues and Methods from a Policy Perspective." In *Measuring Corruption*, edited by Charles Sampford, Arthur Shacklock, Carmel Connors, and Fredrik Galtung, 131–62. Farnham Surrey, UK: Ashgate.
- Earl, Sarah, Fred Carden, and Terry Smutylo. 2001. *Outcome Mapping: Building Learning and Reflection into Development Programs*. Ottawa: International Development Research Centre.
- European Commission. 2004. *Project Cycle Management Guidelines*. Brussels: European Commission EuropeAid Cooperation Office.
- . 2006a. *Cause-and-Effect Analysis*. Brussels: European Commission. [http://ec.europa.eu/europeaid/evaluation/methodology/methods/mth\\_att\\_en.htm](http://ec.europa.eu/europeaid/evaluation/methodology/methods/mth_att_en.htm).
- . 2006b. *Evaluation Methods for the European Union's External Assistance*, vol. 1, *Methodological Bases for Evaluation*. Brussels: European Commission.
- . 2009. *Evalsed: The Resource for the Evaluation of Socio-Economic Development*, Sourcebook 2, *Methods & Techniques—Evaluability Assessment*. Brussels: European Commission. [http://ec.europa.eu/regional\\_policy/sources/docgener/evaluation/evalsed/sourcebooks/method\\_techniques/structuring\\_evaluations/evaluability/index\\_en.htm](http://ec.europa.eu/regional_policy/sources/docgener/evaluation/evalsed/sourcebooks/method_techniques/structuring_evaluations/evaluability/index_en.htm).
- European Partners against Corruption Working Group. 2008. *Common Standards and Best Practices for Anti-Corruption Agencies*. [http://www.stt.lt/documents/tarptautinis\\_bendradarbiavimas/KNAB\\_elektroniskais\\_buklets.pdf](http://www.stt.lt/documents/tarptautinis_bendradarbiavimas/KNAB_elektroniskais_buklets.pdf).
- Foresti, Marta, Bhavna Sharma, Tammie O' Neil, and Alison Evans. 2007. *Evaluation of Citizens' Voice and Accountability: Evaluation Framework*. London: Overseas Development Institute.
- Galtung, Fredrik. 2005. "Measuring the Immeasurable: Boundaries and Functions of (Macro) Corruption Indices." In *Measuring Corruption*, edited by Charles Sampford, Arthur Shacklock, Carmel Connors, and Fredrik Galtung, 101–30. Farnham Surrey, UK: Ashgate.
- Gertler, Paul J., Sebastian Martinez, Patrick Premand, Laura B. Rawlings, and Christel M. J. Vermeersch. 2011. *Impact Evaluation in Practice*. Washington, DC: World Bank.

- Gittinger, J. Price. 1982. *Economic Analysis of Agricultural Projects*. 2nd ed. Baltimore: Johns Hopkins University Press.
- Glasson, John, Riki Therivel, and Andrew Chadwick. 2005. *Introduction to Environmental Impact Assessment*. 3rd ed. London: Routledge.
- ICAC (Independent Commission against Corruption, Hong Kong Special Administrative Region). 2009. *2009 Annual Report*. Hong Kong: ICAC. [http://www.icac.org.hk/filemanager/en/Content\\_1020/2009.pdf](http://www.icac.org.hk/filemanager/en/Content_1020/2009.pdf).
- . 2011. *Performance Pledge*. Hong Kong: ICAC. [http://www.icac.org.hk/filemanager/en/Content\\_1017/pledge\\_eng.pdf](http://www.icac.org.hk/filemanager/en/Content_1017/pledge_eng.pdf).
- IEG (Independent Evaluation Group). 2006. *Impact Evaluation: The Experience of the Independent Evaluation Group of the World Bank*. Washington, DC: World Bank.
- KNAB (Latvia Corruption Prevention and Combating Bureau). 2009. *Progress and Results in Preventing and Combating Corruption in Latvia: Periodical Update*. Riga: KNAB. [http://www.knab.gov.lv/uploads/eng/periodic\\_update\\_20091.pdf](http://www.knab.gov.lv/uploads/eng/periodic_update_20091.pdf).
- Knack, Stephen. 2006. *Measuring Corruption in Eastern Europe and Central Asia: A Critique of the Cross-Country Indicators*. Policy Research Working Paper 3968. Washington, DC: World Bank.
- Langseth, Petter. 2006. "Measuring Corruption." In *Measuring Corruption*, edited by Charles Sampford, Arthur Shacklock, Carmel Connors, and Fredrik Galtung, 7–44. Farnham Surrey, UK: Ashgate.
- Lusthaus, Charles, Marie-Hélène Adrien, Gary Anderson, Fred Carden, and George Plinio Montalván. 2002. *Organizational Assessment: A Framework for Improving Performance*. Ottawa: International Development Research Centre; Washington, DC: Inter-American Development Bank.
- Lusthaus, Charles, Gary Anderson, and Elaine Murphy 1995. *Institutional Assessment: A Framework for Strengthening Organisational Capacity of IDRC's Research Partners*. Ottawa: International Development Research Centre.
- McGee, Rosemary, and John Gaventa. 2010. "Review of Impact and Effectiveness of Transparency and Accountability Initiatives." Paper prepared for Transparency and Accountability Initiative Workshop, October 14–15.
- Meagher, Patrick. 2002. *Anti-Corruption Agencies: A Review of Experience*. Washington, DC: World Bank.
- . 2004. *Anti-Corruption Agencies: A Review of Experience*. IRIS Discussion Paper 04/02. College Park, MD: University of Maryland Center for Institutional Reform and the Informal Sector.
- Morra Imas, Linda G., and Ray C. Rist. 2009. *The Road to Results: Designing and Conducting Effective Development Evaluations*. Washington, DC: World Bank.
- OECD (Organisation for Economic Co-operation and Development). 1996. *Ethics in the Public Service: Current Issues and Practice*. PUMA Occasional Papers. Paris: OECD.
- OECD/DAC (Organisation for Economic Co-operation and Development, Development Assistance Committee). 1986. *Methods and Procedures in Aid Evaluation*. Paris: OECD.
- . 1991. *Principles for Evaluation of Development Assistance*. Paris: OECD.
- . 2000. *Glossary of Evaluation and Results-Based Management (RBM) Terms*. Paris: OECD.
- . 2001. *Evaluation Feedback for Effective Learning and Accountability*. Paris: OECD.
- . 2002. *Glossary of Terms in Evaluation and Results-Based Management*. Paris: OECD.

- . 2006. “DAC Evaluation Quality Standards (for test phase application).” Draft. DAC Network on Development Evaluation, OECD, Paris. <http://siteresources.worldbank.org/EXTGLOREGPARPROG/Resources/DACEvaluationQualityStandards.pdf>.
- . 2007. *Specialised Anti-Corruption Institutions: Review of Models*. Paris: OECD.
- . 2009. *Glossary of Key Terms in Evaluation and Results-Based Management*. Paris: OECD.
- . 2010a. *Evaluating Development Co-operation: Summary of Key Norms and Standards*. 2nd ed. Paris: OECD DAC Network on Development Cooperation.
- . 2010b. *Quality Standards for Development Evaluation*. DAC Guidelines and Reference Series. DAC Network on Development Evaluation. Paris: OECD. <http://www.oecd.org/dataoecd/30/62/36596604.pdf>.
- . 2011. “DAC Criteria for Evaluating Development Assistance.” OECD Development Co-operation Directorate. [http://www.oecd.org/document/22/0,2340,en\\_2649\\_34435\\_2086550\\_1\\_1\\_1\\_1,00.html](http://www.oecd.org/document/22/0,2340,en_2649_34435_2086550_1_1_1_1,00.html).
- Olken, Benjamin A. 2007. “Monitoring Corruption: Evidence from a Field Experiment in Indonesia.” *Journal of Political Economy* 115, no. 2: 200–49.
- Osborne, David, and Ted Gaebler. 1992. *Reinventing Government*. New York: Penguin.
- Picciotto, Robert. 2005. “The Value of Evaluation Standards: A Comparative Assessment.” *Journal of Multidisciplinary Evaluation*, no. 3 (October): 30–59.
- Pope, Jeremy. 1999. *The Need for, and Role of, an Independent Anti-Corruption Agency*. Transparency International Working Paper Series. Berlin: Transparency International.
- Pope, Jeremy, and Frank Vogl. 2000. “Making Anti-Corruption Agencies More Effective.” *Finance and Development* 37, no. 2 (June): 6–9.
- Quah, Jon. 2000. “Accountability and Anticorruption Agencies in the Asia-Pacific Region.” In *Combating Corruption in Asian and Pacific Economies*, 101–24. Manila: Asian Development Bank.
- Ravallion, Martin. 2009. “Evaluating Three Stylised Interventions.” *Journal of Development Effectiveness* 1, no. 3 (September): 227–36.
- Ravindra, Adikeshavalu. 2004. *An Assessment of the Impact of Bangalore Citizen Report Cards on the Performance of Public Agencies*. ECD Working Paper Series 12. Washington, DC: Operations Evaluation Department, World Bank.
- Recovery & Development Consortium. 2010. *DFID Yemen Social Fund for Development: Impact Evaluation*. Kongens Lyngby, Denmark: Recovery & Development Consortium. [http://www.sfd-yemen.org/SFD\\_SITE/admin/PDF/upload\\_DOC/SFD\\_Impact\\_Evaluation\\_Report\\_2010.pdf](http://www.sfd-yemen.org/SFD_SITE/admin/PDF/upload_DOC/SFD_Impact_Evaluation_Report_2010.pdf).
- Reinikka, Ritva, and Jacob Svensson. 2000. *Cost Efficiency in Healthcare*. Washington, DC: World Bank.
- . 2003. *Survey Techniques to Measure and Explain Corruption*. Washington, DC: World Bank.
- Schiavo-Campo, Salvatore. 1999. “Performance in the Public Sector.” *Asian Journal of Political Science* 7, no. 2: 75–87.
- Smilov, Daniel. 2009. “Anti-Corruption Bodies as Discourse-Controlling Instruments: Experiences from South East Europe.” In *Governments, NGOs and Anti-Corruption: The New Integrity Warriors*, edited by Luís De Sousa, Barry Hindess, and Peter Larmour. London: Routledge.
- Søreide, Tina. 2003. *Estimating Corruption*. Bergen, Norway: Chr. Michelsen Institute.
- . 2006. *Is It Wrong to Rank? A Critical Assessment of Corruption Indices*. CMI Working Paper 2006:1. Bergen, Norway: Chr. Michelsen Institute.

- Stake, Robert E. 1995. *The Art of Case Study Research*. Thousand Oaks, CA: Sage.
- Sundet, Geir. 2008. *Following the Money: Do Public Expenditure Tracking Surveys Matter?* U4 Issue 2008:8. Bergen, Norway: U4 Anti-Corruption Centre.
- TI and UNDP (Transparency International and United Nations Development Programme). 2007. *Mapping of Corruption and Governance Measurement Tools in Sub-Saharan Africa*. Berlin: Transparency International; Oslo: UNDP Oslo Governance Centre.
- UNDP (United Nations Development Programme). 2005. *Institutional Arrangements to Combat Corruption: A Comparative Study*. Bangkok: UNDP.
- . 2008. *A User's Guide to Measuring Corruption*. Oslo: UNDP Oslo Governance Centre.
- . 2011. *Practitioners' Guide: Capacity Assessment of Anti-Corruption Agencies (ACAs)*. New York: UNDP.
- UNIFEM (United Nations Development Fund for Women). 2009. *Guidance Note on Carrying Out an Evaluability Assessment*. Evaluation Guidance Note Series. New York: UNIFEM Evaluation Unit.
- University of London. 2010. *Project Appraisal and Impact Analysis Course Reader*. London: Centre for Financial and Management Studies, School of Oriental and African Studies, University of London.
- UNODC (United Nations Office on Drugs and Crime). 2002. *United Nations Global Programme against Corruption Anti-Corruption Toolkit*. Vienna: UNODC.
- USAID (United States Agency for International Development). 2006. *Anti-Corruption Agencies (ACAs)*. Office of Democracy and Governance, Anti-Corruption Programme Brief. Washington, DC: USAID.
- Wade, Robert. 1982. "The System of Administrative and Political Corruption: Canal Irrigation in South India." *Journal of Development Studies* 18, no. 3: 287–328.
- . 1985. "The Market for Public Office: Why the Indian State Is Not Better at Development." *World Development* 13, no. 4: 467–97.
- Webb, Eugene J., Donald T. Campbell, Richard D. Schwartz, and Lee Sechrest. 1966. *Unobtrusive Measures: Nonreactive Measures in the Social Sciences*. Chicago: Rand McNally.
- White, Howard. 2009. *Theory-Based Impact Evaluation: Principles and Practice*. 3ie Working Paper 3. New Delhi: International Initiative for Impact Evaluation.
- Wholey, Joseph S., Harry P. Hatry, and Kathryn E. Newcomer, eds. 2004. *Handbook of Practical Programme Evaluation*. San Francisco: Jossey-Bass.
- Yin, Robert K. 2003. *Case Study Research: Designs and Methods*. Thousand Oaks, CA: Sage.

U4 Anti-Corruption Resource Centre  
Chr. Michelsen Institute (CMI)  
Phone: +47 47 93 80 00  
Fax: +47 47 93 80 01  
u4@u4.no  
www.U4.no

P.O.Box 6033 Bedriftssenteret  
N-5892 Bergen, Norway  
Visiting address:  
Jekteviksbakken 31, Bergen

This U4 Issue is also available at:  
[www.u4.no/publications](http://www.u4.no/publications)

Indexing terms:  
Anti-corruption agencies  
Monitoring and evaluation  
Evaluation standards  
Impact evaluation  
Indicator  
Theory of change  
Quality assurance

Photo:  
Patrick Q  
[patrick\\_q](https://www.flickr.com/photos/patrick_q/) on flickr.com



## Abstract

The number of Anti-corruption agencies (ACAs) around the world has increased dramatically over the past decades. Nevertheless, the value of ACAs is increasingly being questioned by international donors and national governments. Frequently, ACAs are not considered to deliver on the high expectations bestowed upon them.

Evaluations of individual agencies were collected and analysed to assess the evidence underlying the assumptions about the effectiveness of ACAs. Surprisingly, few evaluations had actually been done, and even fewer measured the actual outcomes and impacts of the ACA. Thus, whilst opinions about ACAs are many, the actual evidence about their performance is scarce. To develop this body of evidence, ACAs need to do a better job at establishing results-based indicators for their work, showing how activities lead to impact, and collecting data.

To which extent the perceived failure of ACAs is an issue of measurement or design can therefore not be answered with any certainty. The value of ACAs can only be determined once evidence-based evaluations are conducted.

To this end, the report provides technical, methodological, and practical guidance to assist staff of ACAs in undertaking monitoring and evaluation and shows how the outcomes and impact of the work of ACAs can be evaluated in an objective, evidence-based manner.