

INTOSAI



*Program Evaluation for
SAIs*

A Primer

Juillet 2010

International Organization of Supreme Audit Institutions
Working Group on Program Evaluation

Chair

Mr. Didier Migaud

Premier Président de la Cour des Comptes

Président du Groupe de travail pour l'évaluation de programmes

Cour des Comptes

13 Rue Cambon

F 75100 Paris

tel: ++33 (1) 42 98 95 74

fax: ++33 (1) 42 98 96 02

presidence@ccomptes.fr

Members

Austria • Belgium • Chile • Costa Rica • El Salvador • Finland • Gabon • Georgia •
Germany • Libyan Arab Jamahiriya • Lithuania • Mexico • Morocco • Pakistan •
Papua New Guinea • Poland • Republic of Korea • Switzerland • United States of
America

*Netherlands and Sweden, though not belonging to the group any longer,
participated actively to its work in the past years.*

*F*oreword



**Foreword by Didier Migaud, First President of the French *Cour des Comptes*,
President of the INTOSAI Working Group on Program Evaluation**

In 1992, the 14th INTOSAI Congress held in Washington, D.C. established the Working Group on Program Evaluation and entrusted its presidency to the French Cour des Comptes. Evaluation seemed to be a still new but promising initiative which could enhance the audit practices used by the supreme audit institutions.

With the active participation of its 20 members and through the contribution of all the SAIs who responded to the questionnaires, our Working Group was able to gain a more in-depth knowledge of evaluation: How to define and position it with respect to performance audits, what are the principal stages and the terms of its implementation. Each of the INTOSAI Congresses at which we have reported on our activities has confirmed the mandate of our Group: making access to evaluation

easier for SAIs and clarifying the concepts, methods, and practices of evaluation.

Evaluation is sharply different from our traditional activities, including that of performance audits, which is the area that seems closest to evaluation. Evaluation extends and supplements the performance audit as the only tool that makes it possible to grasp complex public policies, to explain why such policies have succeeded or why they have failed, to identify their objectives and to understand their implementation.

This is why the practice of evaluation compels us to make adjustments in our organization, our methods, and in our very culture. The evaluation of complex public actions requires rigorous methods, listening to stakeholders, and the increasing use of pluralist expertise. That is the price at which we will meet the expectations of our fellow citizens, who would like public policies to be both more readable and more effective.

This document, which has been prepared with the invaluable support of the United States GAO, is based on all the work performed by the Working Group since its creation. This is not a collection of standards nor is it a practical guide, even if it offers numerous concrete illustration of the Evaluation Program of the SAIs. It was produced to make it easier to get started with evaluation, to map out its course, and to provide its main points of reference. From that point on, it is up to everyone to trace their own path and to find, for each evaluation, the methods that suit them best.

There is an open road ahead of us: It is now up to us to take it and invite as many as possible to join us. If we want evaluation to grow within the SAIs, we must foster capitalizing on knowledge and the spreading of good practices. This is the program that our Working Group has set for itself for the coming years, a program that is in line with the INTOSAI strategic direction. I am counting on all of you to carry this through to a successful conclusion!

Didier Migaud

Contents

Foreword	i
Contents	v
Introduction	ix
Chapter 1: The Challenge of Program Evaluation	1
[1.1] New Perspectives on Performance	1
[1.2] Thinking about Strategies for Change	9
[1.3] The Value in Evaluation	13
Notes	14
Chapter 2: Defining Program Evaluation	16
[2.1] SAIs Variously Define Financial and Performance Audits	16
[2.2] Some Institutions Define Performance Audit Generically to Include Program Evaluation	16
[2.3] Other International Organizations Define Evaluation	16
[2.4] An Operational Definition of Program Evaluation for SAIs	16

Notes 16

Chapter 3: Planning for Program Evaluation 16

[3.1] Working with Evaluation’s Stakeholders 16

[3.2] Adopting a Design Strategy 16

[3.3] Linking Design to Evaluation Questions 16

[3.4] Assessing a Team’s Capacity to Execute Evaluation Plans 16

Notes 16

Chapter 4: Conducting Program Evaluation 16

[4.1] Working through Evaluation Methodology 16

[4.1.a] Some Directly Available Methodology Resources 16

[4.1.b] Some Indirectly Available Methodology Resources 16

[4.2] Framing Project Management Plans 16

Notes 16

Chapter 5: Developing Organizational Capacity for Program Evaluation 16

[5.1] The Evaluation Team 16

[5.2] Developing Organizational Expertise 16

[5.3] The Evaluation Community 16

Notas 16

Chapter 6: Perspectives on the Future..... 16

[6.1] SAIs as Agents of Change 16

[6.2] Pathways to New Learning 16

[6.3] Supporting an Evaluation Culture 16

Notes 16

Appendix A: Evaluators' Resources 16

Appendix B: References..... 16

Appendix C: Glossary 16

Index of Studies 16

Abbreviations

AEA	American Evaluation Association
AfrEA	African Evaluation Association
ANAO	Australian National Audit Office
Danida	Danish International Development Assistance
EEA	European Evaluation Association
EU	European Union
GAO	U.S. General Accounting Office
INTOSAI	International Organization of Supreme Audit Institutions
OECD	Organisation for Economic Co - operation and Development
SAI	Supreme Audit Institution
USAID	U.S. Agency for International Development

*I*ntroduction

Program Evaluation for SAIs: A Primer introduces Supreme Audit Institutions (SAI) to ways in which they may make evaluation “an integral part of [their] organizations’ everyday operations . . . part of the responsibilities.”¹ At the final session of the American Evaluation Association’s 16th annual conference in 2001, it was suggested that rather than mandating evaluation, evaluators should tell people that

evaluation is hard and possibly really profitable but we [evaluators] want to involve only those who are willing to try. . . . “Tell them everyone can’t do it! Only those who really want to demonstrate excellence should get involved. Use reverse psychology!”²

Later in the session, it was concluded that

the fact that not everyone agrees on what evaluation is isn’t a problem. “The strength of our field is its diversity. Even if we agreed on the term evaluation there are lots of ways to do that. We’ve demonstrated that we won’t agree. . . . Evaluative thinking, ways of being explicit and bringing evidence to bear, runs through all kinds of evaluation. The challenge is in the application to different situations.”³

At the end of the session, James R. Sanders wrapped up the conference this way:

I continue to look at mainstreaming as adopting evaluation as a core value of an organization or a professional practice. Evaluation is becoming part of the culture, a common language. . . . The question about what we [evaluators] are asking people to mainstream is a critical question. To me we are asking people to mainstream a frame of mind, to ask good questions, be skeptical, use answers to bring about change, to continue to move toward excellence in

practice. . . . As evaluators, we ought to always be evaluating our evaluations That is part of mainstreaming—asking good questions about our own practice⁴.

SAIs may think it somewhat contrary to be introduced to a paper on program evaluation with remarks from the closing of a conference on evaluation. And yet mainstreaming—“*building an evaluation culture* throughout our programs, organizations, and policy development [even though] such evaluative activity may not necessarily be labeled evaluation”—is what *Program Evaluation for SAIs: A Primer is all about*⁵.

Supreme Audit Institutions abound with reports of financial audits, value - for - money audits, sectoral audits, performance audits, program effectiveness audits, and publications in various categories, many of which on close reading appear to be quite evaluative, whether explicitly or not. Without making any representation as to being evaluative in any way, for example, the Audit Office for Wales’ strategic plan for 2004–2009 set forth a program that is clearly achievable only by evaluative means, as the six chapters of this primer will show:

The Wales Audit Office is able to maximise its beneficial impact on public services over and above that of its predecessor organisations by exploiting its ability to . . . interact with and influence a wider range of stakeholders across whole systems and through greater cross - sector working.

There are a number of circumstance and opportunities that will assist the Wales Audit Office to achieve its potential.

These include . . . support and regard for our role in promoting improvement by stakeholders in all Welsh public services [and] . . . our ability to look across the whole of public services in Wales

[T]he Wales Audit Office has adopted a matrix approach to planning and delivering its services which better matches people and skills to projects. We will make wider use of our range of methodologies. We will broaden our staff’s experience and encourage whole systems thinking through matrix

management We will develop . . . methods for evaluating the outcomes of partnerships and collaborative working. We will ensure our work is focused on the user. We will examine the role of the voluntary sector, community and private sector when evaluating whole systems....⁶

What does program evaluation demand of SAIs? It demands thinking about strategies for change, for becoming more evaluative, within the organization and the organization's culture. It demands learning new competencies, such as how to design and plan and execute evaluations and how to work with the people who request them. It demands a new kind of learning in terms of knowing how to look for data sources and methods of analysis and how to ensure their quality. It means building a capacity for evaluation, not only so that individual staff and staff working in evaluation teams can grow in knowledge, skill, and ability but also so that the organization will increase its evaluative capacity. Program evaluation demands new perspectives on management planning. And it especially demands a tolerance for ways of working and thinking that may seem to break with convention. It expects that SAIs will look to the future and undertake new pathways to learning and practice.

The chapters of this paper walk SAIs through some of the basic concepts of program evaluation, which many have already been practicing—some implicitly, some explicitly. Many examples throughout the text—as well as in the textboxes—support and amplify these concepts. All are taken directly from national audit institutions, professional evaluation associations, and other professional evaluation groups. They range from Bangladesh to New Zealand, from France to the People's Republic of China, from Georgia to Brazil.

Chapter 1 discusses some of the inter - relationships between performance audit and program evaluation and some of the many possibilities for SAIs to become more evaluative through the development of crossover practices between them. The scope of work in performance audits may be narrow or broad. Performance audits may apply conventional auditing guidelines or a variety of other methodologies, including social science methods. And they may involve various levels of research, analysis, or evaluation. SAIs need no special mandate to conduct program evaluation.

Chapter 2 shows a few of the different ways SAIs around the world use financial audit, performance audit, and program evaluation and terms related to them. Some definitions are mandated by nations' laws; others are the result of variable practice. Literature reviews attempting to summarize how program evaluation has been defined in different places and over time have discovered that it has no one universal denotation, either overall or when divided into its elements.

At the end of chapter 2, an operational definition of program evaluation is offered—one in which SAIs may find sufficient and necessary elements from which to build a foundation for their own best definition. For example, networking members of the African Evaluation Association might wish to help adapt the definition to SAI needs, since AfrEA has expressed some concern about “the relevance of the U.S. ‘Program Evaluation Standards’ to evaluation work in Africa” in its *African Evaluation Guidelines 2000: A Checklist to Assist Evaluators*.

Chapters 3 and 4 resume the discussion of program evaluation's challenge. Chapter 3 presents some of the challenges of planning and designing evaluations, given the constraints evaluators operate under when they attempt to tell the public a credible performance story about policies, programs, and other entities.

Chapter 4 points SAIs toward resources for finding the methods they need to execute well - designed evaluations. The community of evaluators—within evaluative institutions and associations of professional evaluators— provides a wealth of “checklists,” “cook - books,” “matrixes,” and various other data sources and analysis methods. These are widely available, in print and electronically, for help in matching appropriate methodologies to evaluation design questions. Chapter 4 is a starting place.

Chapter 5 alerts SAIs to some of the new ways of working that program evaluation entails. Developing the capacity for program evaluation means developing new staff competencies and new methods within the institution. It also means moving outward to engage with the community of external evaluation experts who work in academia, as consultants, and in professional evaluation associations around the world.

Finally, chapter 6 offers a perspective on the future: implications for strategic

relationships between audit and evaluation societies, between INTOSAI work groups, and between public and private institutions.

The paper concludes with three appendixes. Appendix A, “Evaluators’ Resources,” offers links to a handful of international organizations that make available fellowships, study grants, coursework, professional meetings, job opportunities, and other resources for developing program evaluation skills. Appendix B is a list of references, and appendix C is a glossary of terms related to evaluation that are employed in this paper.

To establish a program evaluation unit within an ongoing organization will require concentrated effort and a long - term commitment from dedicated managers and experienced evaluators—and some decisions about risk. To foster an evaluation culture will require working, over time, to put in place a formal, regular process for planning, executing, and using information from evaluations, for making a commitment to learning through analysis and experiment. To identify what’s uniquely best for the individual organization within the growing network of evaluating organizations around the world will take nothing less than hard work.

Readers who look to these pages for an easy formula for organizational or cultural change will not find one. This paper is neither a textbook nor a workbook. It is an outline of concepts, examples, and resources. It suggests methods, pathways, and possibilities. For some, it tells stories they already know. For others, it offers new ideas, tools, and support for change and growth.

1 See James R. Sanders, “Mainstreaming Evaluation,” in J. Jackson Barnette and James R. Sanders, eds., *The Mainstreaming of Evaluation*, New Directions for Evaluation No. 99 (San Francisco, Calif.: Jossey - Bass, 2003), p. 3. Sanders was speaking at the American Evaluation Association’s 2001 conference: see American Evaluation Association, *Evaluation 2001*, November 7–10, St. Louis, Missouri. <http://www.eval.org/Training/conferencehistory.asp> (Apr. 7, 2009).

2 David D. Williams and Mark L. Hawkes, quoting M. Q. Patton, in “Issues and Practices Related to Mainstreaming Evaluation: Where Do We Flow from Here?” in Barnette and

Sanders, *The Mainstreaming of Evaluation*, p. 77.

- 3 Williams and Hawkes, quoting Patton, in Barnette and Sanders, *The Mainstreaming of Evaluation*, p. 77.
- 4 Williams and Hawkes, quoting Sanders, in Barnette and Sanders, *The Mainstreaming of Evaluation*, p. 79. Sanders served 26 years as associate director of the Evaluation Center at Western Michigan University until he retired in 2001, chaired the Joint Committee on Standards for Educational Evaluation 1988 to 1998, and in 2001 was president of AEA.
- 5 Paul Dignan, “Mainstreaming Evaluation or Building Evaluation Capability? Three Key Elements,” in Barnette and Sanders, *The Mainstreaming of Evaluation*, p. 12. Emphasis in original.
- 6 Auditor General for Wales, *Making Public Money Count: Wales Audit Office Strategy 2006–2009* (Cardiff: Wales Audit Office, May 2006), pp. 3–4 and 9–10. Wales Audit Office, Publications, Our Plans and Accounts. www.wao.gov.uk (Apr. 7, 2009).

Chapter 1: The Challenge of Program Evaluation

This paper's six chapters challenge Supreme Audit Institutions (SAI) and other public managers to look in a new way at program evaluation, often subsumed under performance audits. Canada's Office of the Auditor General issued a discussion paper in spring 2003, *Reporting on Outcomes: Setting Performance Expectations and Telling Performance Stories*, suggesting that "Reporting on performance . . . involves telling a credible *performance story* . . ." ¹ The chapters of this paper tell about the value that program evaluation has added to some countries' reporting function and the additional value it may have for countries building a capacity for greater evaluative skill among staff and within the larger evaluation community.

[1.1] New Perspectives on Performance

Taking this new perspective, INTOSAI has noted recently that

a common objective of most governments today is to improve the quality of public services, particularly as people's expectations (often with reference to the service they receive from the private sector) of what constitutes quality continue to increase. To promote improvements of this type, many governments have embarked on modernisation programs to deliver better services that are, for instance, more easily accessible and convenient, provide citizens with more choice, and are delivered more quickly. The quality of public services is an increasingly important issue, which members of parliaments and governments across the world expect the SAIs to address in

their performance audit reports ².

[1] VALUE IN EVALUATION:

[W]hat does “good” mean anyway . . . ? As Wittgenstein suggested, “good” . . . has a family of meanings. Prominent among them is this one: “meets the criteria or standards of . . . evaluation.”

John R. Searle, *Speech Acts: An Essay in the Philosophy of Language* (New York: Cambridge University Press, 1969), p.

SAIs may be seen, if a hypothetical survey were run, as a group usually conducting financial audits, oriented toward some combination of accountability, compliance, control, and correction (Brazil and Greece, among others); a group also conducting performance audits, focusing on effectiveness as well as economy and efficiency (Korea and Sweden, for example); and a group that often goes beyond performance audits toward assessing public policy and evaluating public programs (France and the United States).³ More formally, program evaluation’s role for SAIs has been noted in the 2004 document entitled *Implementation Guidelines for Performance Auditing: Standards and Guidelines for Performance Auditing Based on INTOSAI’s Auditing Standards and Practical Experience*:

In recent years, the concept of program evaluation has been a growing subject of discussion amongst SAIs. Whether or not program evaluation is an important task for a SAI has been discussed. A special group (INTOSAI Working Group on Program Evaluation) has been set up to promote principles and guidance in this area. It is generally accepted that program evaluation has objectives identical or similar to those of performance auditing in that it seeks to analyse the relationship between the objectives, resources, and results of a policy or program. It has also been agreed that program evaluation is an important task for a SAI that has the authority and competence to carry out such studies.⁴

It would seem that SAIs are increasingly reshaping their methods in ways implied by the *Implementation Guidelines for Performance Auditing*. Those following such

trends include, for example, the Audit Bureau of Jordan, which has detailed, within its strategic plan, long - term goals for 1998–2008 of “abandoning the pre - audits [and] . . . intensifying the work on performance auditing,” noting that among its functions is suggesting “methods that contribute to managing resources in an economic, efficient, and effective way.”⁵

SAI projects today can run the gamut from evaluations that remain listed among countries’ audit reports to evaluations that are given their own category in countries’ publications lists and that fully credit the projects’ analysts as program evaluators. Many who conduct performance audits, following conventional auditing guidelines, are becoming more evaluative in what they do, applying and adopting broader methods, including social science methods, as they examine relationships between a program’s resources and objectives, for example, or a policy’s objectives and results. Consider, for example, a February 19, 2009, report from the Netherlands entitled *Comparative Study of the Reports of Supervisory Boards: Good Examples and Tips for Better Reports*. The Court of Audit staff who developed the project appeared to have been as concerned with program effects as with performance. They found that

“Good governance” has been high on the agenda of organisations in both the public and the private sectors in recent years. In the public sector it has led to the development of a wide range of good governance codes. Good internal supervision of management is one of the key principles of good governance...

Supervisory boards are a relatively new phenomenon in the public sector and their definite form still has to emerge in practice. The supervisory boards’ accountability for their own performance is also still developing. There is a great deal of variation in the way in which they render account.

...

The importance of supervisory boards and their public accountability to politicians and society prompted us to take a look at the reports they issue. We investigated 84 reports (on the 2007 reporting year) and websites of supervisory boards at care institutions, educational institutions and

autonomous administrative authorities. We compared them with the good governance codes applicable in the respective sectors and with relevant standards applied by the Court of Audit. By analysing the reports, we wanted to gain an insight into the information value of the supervisory boards' reports and contribute to the further development of the phenomenon of supervisory boards as links in the supervisory chain.⁶

Canada provides another example of performance audit crossing over toward program evaluation. In 2001's *Public Service Management Reform: Progress, Setbacks and Challenges*, the Auditor General of Canada reported that

In 1979, the Royal Commission on Financial Management and Accountability (the Lambert Commission) noted that the management of personnel in all its aspects is as important as, if not more important than, financial management in achieving overall management of government activities; and yet Parliament's review of personnel management was more limited than its review of financial management matters through the Public Accounts Committee. This remains the case today.⁷

Undertaking a study, therefore,

of the efforts at public service reform over the last decade to provide Parliament with a broad assessment of their progress, a perspective on the challenges the government and the public service now face in moving forward with needed change, and our views on prerequisites for success,⁸

his office found that despite those reform efforts,

Substantial improvements are still necessary in areas such as modernizing and improving service to the public and making better use of performance information to strengthen programs, achieve desired results, and account for performance.⁹

The study's methodology was to review "a mass of documentation, including the major employee survey conducted in 1999," to conduct interviews

with 13 subject matter specialists in central agencies and with a variety of others [including] . . . a dozen executives in the National Capital Region, . . . eight regional heads, . . . and four people who were senior union officials during all or much of the decade,

and, to offset the difficulty of respondents' recall, to analyze "statistical information indicating the scale of change."¹⁰

Even a project in Finland that monitored product safety might be interpreted as crossing over indirectly from audit into evaluation:

Product safety monitoring has not been supervised in an active and thorough way. . . . One indication of the passive approach to supervision is that local authorities have not been aware of key principles which the Consumer Agency considers important in product safety monitoring at the local level. . . . Monitoring has not been proportional to municipalities' size, industrial structure or risks. Some small rural municipalities have invested much more in monitoring than some large cities.

A direct evaluation of the effectiveness of monitoring cannot be presented on the basis of the audit. State supervising authorities have considered effectiveness to depend largely on monitoring by local authorities. . . . The Consumer Agency believes that some home and leisure accidents have been caused by goods and services which do not comply with regulations and guidelines and that even limited monitoring has to some extent prevented health risks, accidents and even fatalities. Detailed monitoring information is not available, however.¹¹

A last example from an SAI report is from Brazil, although many more may be found internationally across the Internet—and perhaps in the archives of SAIs not yet online. In a 2003 report entitled *TCU Evaluation of the Actions for Detection and Correction of Visual Impairment*, staff of the Brazilian Court of Audit stated that they had set out with the following objective:

Considering that access to education is a constitutional right, the audit proposed to evaluate the extent to which the PNSE [National Program of

Schoolchildren's Health] has been contributing to improve the learning process of the school population with visual impairment. This was done by analyzing five questions:

- 1 How does the Campaign ["Look Me in the Eye"] contribute to the improvement of the school performance of its beneficiaries?
- 2 Is the criterion for selection of municipalities to be covered by the 2001 Campaign the one that best meets the principle of equity?
- 3 Was there loss of students in the various stages of the 2000 Campaign?
- 4 Was the delivery of the spectacles to the students timely in the 2000 Campaign?
- 5 Would new partnerships increase access to the Campaign? ¹²

As chapter 3 will show, these are typically questions a program evaluation team might frame in attempting to meet the objectives of an evaluation rather than those of an audit.

In addition to reporting the evaluation's purpose, scope and methodology, findings, and recommendations, evaluators stated benefits that implementing the Brazilian Court of Audit's recommendations would bring to the children of Brazil and their communities, both quantitatively and qualitatively.

Several of these benefits would derive from considerable change to the program—including modifying its timetables; implementing a computer system; disseminating information to the states, municipalities, communities, families, and the press; and establishing health care partnerships.¹³

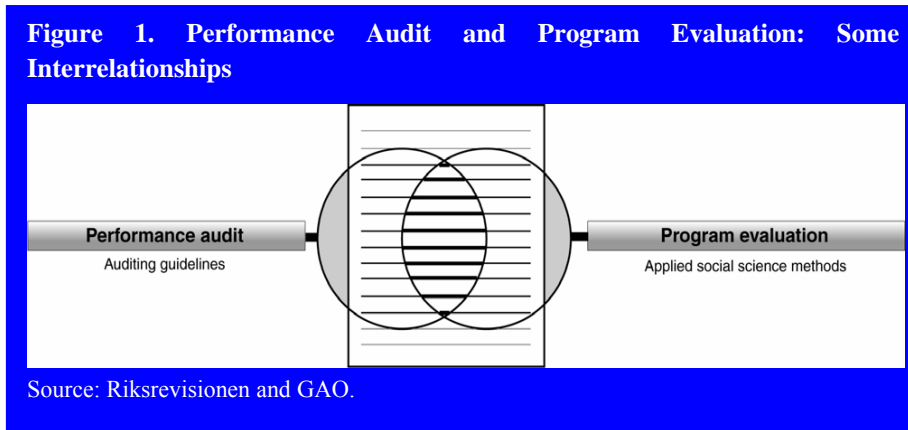
INTOSAI's *Implementation Guidelines for Performance Auditing* points out that

A number of SAIs are not required to execute performance audits or may consider themselves somewhat limited in their capacity and experience in respect of these audits. Other SAIs may have a long tradition of carrying out

both advanced performance audits and complex program evaluations. . . .¹⁴

Performance audit and program evaluation are often interrelated. The scope of work in performance audits may be narrow or broad.

Figure 1 suggests a model of the many possibilities for SAIs' becoming more evaluative through the development of crossover practices. In performance audits they may apply conventional auditing guidelines (as seen on the left of the figure) or a variety of other methodologies, including social science methods (as seen on the right). And they may involve various levels of research, analysis, or evaluation.



SAIs need no special mandate to conduct program evaluation. Any SAI who wishes to may perform evaluations separately or embed them within a mandate for performance audits (the center of figure 1). Whichever course SAIs take may lead to organizational change and may even strengthen the organization.

The Republic of the Philippines, noting that in 2002 it had changed the name of its Special Audit Office to Management Services, listed this office's functions as conducting, among other things, "sectoral performance audits."

The Philippines' Commission on Audit stated on its Web site in 2009 that a sectoral audit

refers to an audit of programs or activities that are delivered by more than one government agency and is expected to provide:

- an overall picture of how various segments of a program are implemented and possibly lead to the identification of areas where improvements can be introduced; . . .
- [a] basis for auditors to realize that program difficulties may not lie with a single agency but possibly with the way the agencies involved in the program work together; . . .
- [an] opportunity for making changes in the program, if necessary.¹⁵

INTOSAI's *Implementation Guidelines for Performance Auditing* echoes these kinds of intentions when it notes two approaches to performance auditing:

The results - oriented approach deals mainly with questions such as: "What is the performance or what results have been achieved, and have the requirements or the objectives been met?" In this approach, the auditor studies performance (concerning economy, efficiency, and effectiveness) and relates observations to the given norms (goals, objectives, regulations, etc.) or the audit criteria (more or less precisely defined before the main study begins). . . .

The problem - oriented approach . . . deals primarily with problem verification and problem analysis, normally without reference to predefined audit criteria. . . . A major task . . . is to verify the existence of stated problems and to analyse their causes from different perspectives (problems related to economy, efficiency, and effectiveness of government undertakings or programs). The problem - oriented approach deals with questions such as: "Do the stated problems really exist and, if so, how can they be understood and what are the causes?"¹⁶

Figure 1 brings both these approaches into focus. The results - oriented approach is

often seen through the lens of performance audit, while the problem - oriented approach may cross over into areas of program evaluation.

[1.2] Thinking about Strategies for Change

It has been said, in the relevant literature, that changing the way an organization perceives and performs its mandate or mission often entails changing the organization's culture. One research study, for example, noted that such

change is a long - term effort that takes at least 5 to 10 years to complete. Company officials believe that two key techniques are of prime importance to a successful culture change:

- Top management must be totally committed to the change in both words and actions.
- Organizations must provide training that promotes and develops skills related to their desired values and beliefs.¹⁷

The study found that greatly important to an organization's ability to change are the techniques it uses to promote values and beliefs: developing a statement of values and beliefs; communicating them to employees; using a management style compatible with values and beliefs; offering rewards, incentives, and promotions to encourage behavior compatible with them; conveying and supporting values and beliefs at organizational gatherings; and making the organization's structure compatible with them.¹⁸ Studies looking specifically at change within SAIs, particularly their modernization of audit practices, have come to somewhat similar conclusions. For example, from Thailand, where "a new budget system [has directed] the State Audit Office to put more emphasis on performance audits," has come a paper entitled Strategy to Improve Planning for Performance Audits that found that

The new organizational structure creates significant change in the Office: the movement of skilled staff members away from their areas of expertise, staff

shortages, etc. [and that]

The heavy demand for Performance Audits has created a series of problems:

- difficulties in selecting and scoping audits
- insufficient evidence to substantiate audit reports
- difficulties in reaching consensus among staff concerning such issues as
- resources, budgets and timeframes and
- a lack of resources, time and budget to complete all audits.¹⁹

The paper recommended that the best strategy to

[2] SAIs' MANY TYPES OF AUDIT AND EVALUATION:

In the area of performance measurement both financial and performance auditors might be involved. In some countries, an individual performance audit may include many different kinds of studies and even several program evaluations. In that sense program evaluation may be considered one of many possible "tools" that performance auditing uses. Program evaluation is one type of study that might be executed by the SAI under the general heading of performance auditing. It is an activity of increasing interest and importance.

INTOSAI Auditing Standards Committee, Implementation Guidelines for Performance Auditing: Standards and Guidelines for Performance Auditing Based on INTOSAI's Auditing Standards and Practical Experience (Stockholm: July 2004), p. 25.

improve the planning process and support the activities of the State Audit office of Thailand [is to]:

- 3.1 Have a clear understanding of the planning process and key players' roles
- 3.2 Using consultation and advisory services
- 3.3 Using risk-based long-range entity planning
- 3.4 Developing planning guidelines for performance audits
- 3.5 Developing monitoring tools
- 3.6 Upgrading communication functions
- 3.7 Supporting continuous learning.²⁰

In Estonia, also, the National Audit Office has placed on its Web site a 2002 document entitled "Measuring Performance Audit Effectiveness: The Case of Estonia," whose point of departure is that

The performance audit has already spread into the practice of many Supreme Audit Institutions. At the same time it has raised discussions about the independence of the auditors and of the more general role of the public sector audit. More and more often, especially after extensive programs of public management reform have been introduced all over the world, the audit institutions have been forced to justify their own activities and demonstrate their results and achievements. Therefore, the ways of assessing the effectiveness of performance audit has gained more importance and attention.²¹

[3] THE COMMON GOOD—CONTROL VS. EVALUATION:

[C]ontrol is essential to administration; without it we cannot avoid straying from the planned results.

To control is . . . above all, to prevent, detect, correct, to guide. . . .

With regard to . . . the agencies and institutions [administrators] manage, the Court [of Audit] will act . . . as a partner who wishes to reach the supreme goal of every public administrator: promote common good.

. . .

Performance audit is the modality of control that has been most developed in the past decades and it had an important role in modernization of public institutions in several countries. The purpose of program evaluation is to examine the effectiveness of government programs and projects. This modality of audit aims at verifying to what extent the actions implemented produced the effects intended by the administration.

TCU has carried out performance audits in several fields, such as: agriculture, education, social assistance, and health. They are considered priority programs . . . aimed at combating poverty and promoting citizenship and social inclusion.

The government program evaluations conducted by the Court in 2002 verified how the Federal Government is investing R\$5.7 billion in a population of beneficiaries made up of over 80 million Brazilians.

Brazil, Brazilian Court of Audit, TCU Annual Report, Fiscal Year 2002 (Brasilia: Tribunal de Contas da União, Planning and Management Secretariat, 2003), pp. 3 and 36–37. <http://www.tcu.gov.br> (Apr. 17, 2009).

Much as with the capacity for program evaluation, the document's author explains, this is owing in no small part to the fact that

A performance audit is a blend of both a conventional audit and management consulting incorporating the knowledge and techniques of many disciplines It benefits from the independence, objectivity, and reporting skills of the auditors, complemented by the specialized analytical systems and implementation skills, that may be available from management consultants .

...

There is no single timeless methodology or set of practices that can be used to adequately conduct a performance audit. In order to be successful an auditor must select those methodologies and techniques, that are appropriate considering the resources available, the nature of the activity chosen and the object of the audit.²²

Examining such findings and those of other studies, along with the studies' recommendations, can be helpful to organizations taking on the challenge of adding program evaluation to performance audit. For example, one report told this story of how agencies had developed their capacity to do program evaluation:

First, agency managers sustained a commitment to accountability and to improving program performance—to institutionalize an evaluation culture. Second, they improved administrative systems or turned to special data collections to obtain better quality data. Third, they sought out—through external sources or development of staff—whatever expertise was needed to ensure the credibility of analyses and conclusions. Finally, to leverage their evaluation resources and expertise, agencies engaged in collaborations or actively educated and solicited the support and involvement of their program partners and stakeholders.²³

[1.3] The Value in Evaluation

As chapter 2 will show, SAIs define financial audit, performance audit, and program

evaluation in a variety of ways. Some definitions are mandated by law; others stem from variable practice. Literature reviews summarizing how program evaluation has been defined in different places and over time have discovered that it has no universal denotation. For example, the *Canadian Evaluation Society Project in Support of Advocacy and Professional Development Evaluation Benefits, Outputs, and Knowledge Elements* reported in 2002 that

- No single [evaluation] output or process is going to be appropriate for all situations and all purposes. The literature does not provide much guidance in this area
- It is hard to separate [program] outputs from benefits, and it is also hard to separate outputs from knowledge/skills.²⁴

The operational definition of program evaluation proposed at the end of chapter 2 may, then, contain sufficient and necessary elements for SAIs to construct their own best definition.

-
- 1 John Mayne, Office of the Auditor General of Canada, *Reporting on Outcomes: Setting Performance Expectations and Telling Performance Stories* (Ottawa, Ontario: May 14, 2003), p. 13 (emphasis in original). www.oag-bvg.gc.ca/internet/English/meth_lp_e_907.html (Apr. 7, 2009).
 - 2 INTOSAI Auditing Standards Committee, *Implementation Guidelines for Performance Auditing: Standards and Guidelines for Performance Auditing Based on INTOSAI's Auditing Standards and Practical Experience* (Stockholm, Sweden: July 2004), p. 21. <http://intosai.connexcc-hosting.net/blueline/upload/1implgperfaude.pdf> (Apr. 7, 2009).
 - 3 For the first group, see, for example, Brazilian Court of Audit, *TCU Annual Report, Fiscal Year 2002* (Brasilia: Tribunal de Contas da União, Planning and Management Secretariat, 2003), p. 5. www.tcu.gov.br (Apr. 17, 2009), and Greece, Supreme Court of Audit, Athens. <http://www.elsyn.gr/elsyn/root.jsp> (Apr. 17, 2009). For the second group, see, for example, Board of Audit and Inspection of Korea, Responsibilities and Functions, Board of Audit and Inspection Act, Seoul, Korea. <http://www.bai.go.kr> (Apr. 17, 2009), and Riksrevisionen, *The Swedish National Audit Office: Booklet 2008*

- (Stockholm: n.d.), p. 2. www.riksrevisionen.se/templib/pages/NormalWithIntrosPage_881.aspx (Apr. 17, 2009). For the third group, see, for example, France, Cour des Comptes, Juridictions financiers, Publications, Rapports publics thématiques, *La formation professionnelle tout au long de la vie* (Paris: October 2008). www.ccomptes.fr/fr/JF/Accueil.html (Apr. 17, 2009), and GAO, Technology Assessment: *Using Biometrics for Border Security*, GAO - 03 - 174 (Washington, D.C.: Nov. 14, 2002). www.gao.gov (Apr. 17, 2009).
- 4 INTOSAI Auditing Standards Committee, *Implementation Guidelines for Performance Auditing*, p. 23.
 - 5 ARABOSAI, ISC members, Jordan, Strategic Plan. www.arabosai.org/en/index.asp (Apr. 17, 2009).
 - 6 Netherlands Court of Audit, Publications, Audits, *Comparative Study of the Reports of Supervisory Boards: Good Examples and Tips for Better Reports* (The Hague, The Netherlands: Feb. 19, 2009). www.rekenkamer.nl (Apr. 17, 2009).
 - 7 L. Denis Desautels, Auditor General of Canada, *Public Service Management Reform: Progress, Setbacks and Challenges* (Ottawa, Ontario: Office of the Auditor General of Canada, Feb. 2001), p. 28. See Office of the Auditor General of Canada, Reports and Publications, Audit Methodology, Studies and Tools. www.oag-bvg.gc.ca/internet/English/admin_e_41.html (Apr. 17, 2009).
 - 8 L. Denis Desautels, Auditor General of Canada, *Public Service Management Reform: Progress, Setbacks and Challenges*, p. 25.
 - 9 L. Denis Desautels, Auditor General of Canada, *Public Service Management Reform: Progress, Setbacks and Challenges*, p. 25.
 - 10 John Edwards, “Looking Back from 2000 at Public Service 2000,” p. 7, in L. Denis Desautels, Auditor General of Canada, *Public Service Management Reform: Progress, Setbacks and Challenges*.
 - 11 National Audit Office of Finland, Publications, Performance Audit Reports, *Product Safety Monitoring*, abstract 69/2004 (Helsinki: 2004), pp. 2– 3. www.vtv.fi (Apr. 17, 2009).

- 12 Brazil, Brazilian Court of Audit, Publicações institucionais, Em inglês, Executive Summaries 6, *TCU Evaluation of the Actions for Detection and Correction of Visual Impairment* (Avaliação do TCU sobre as ações de detecção e correção de problemas visuais) (Brasília: Tribunal de Contas da União, Government Programs Control and Evaluation Secretariat, 2003), executive summary, pp. 8–9. www.tcu.gov.br (Apr. 17, 2009).
- 13 Brazil, Brazilian Court of Audit, *TCU Evaluation of the Actions for Detection and Correction of Visual Impairment*, p. 17.
- 14 INTOSAI Auditing Standards Committee, *Implementation Guidelines for Performance Auditing*, p. 25.
- 15 Republic of the Philippines, Commission on Audit, Government-wide and Sectoral Performance Audit Reports, Quezon City. www.coa.gov.ph (Apr. 17, 2009).
- 16 INTOSAI Auditing Standards Committee, *Implementation Guidelines for Performance Auditing*, pp. 26–27.
- 17 GAO, *Organizational Culture: Techniques Companies Use to Perpetuate or Change Beliefs and Values*, GAO/NSIAD-92-105 (Washington, D.C.: Feb. 27, 1992), p. 2. www.gao.gov (Apr. 17, 2009).
- 18 This list of techniques has been adapted from GAO, *Organizational Culture*, table 1, p. 3.
- 19 Jaruwan Ruangswadipong, *Strategy to Improve Planning for Performance Audits*, Thailand CCAFFCVI Fellow 2000–01 Report, p. 5. www.ccaffcv.com/fellows/reports/Thailand2000.pdf (Apr. 17, 2009).
- 20 Jaruwan Ruangswadipong, *Strategy to Improve Planning for Performance Audits*, p. 5.
- 21 Jane Etverk, “Measuring Performance Audit Effectiveness: The Case of Estonia,” M.A. thesis, Department of Public Administration, University of Tartu, Tartu, Estonia, June 27, 2002, p. 5. National Audit Office of Estonia, About NAO, Documents, Other Documents, Tallinn, Estonia. www.riigikontroll.ee/?lang=en (Apr. 17, 2009).

- 22 Jane Etverk, "Measuring Performance Audit Effectiveness," p. 18.
- 23 GAO, *Program Evaluation: An Evaluation Culture and Collaborative Partnerships Help Build Agency Capacity*, GAO-03-454 (Washington, D.C.: May 2, 2003), p. 14. www.gao.gov (Apr. 17, 2009).
- 24 Rochelle Zorzi, Martha McGuire y Burt Perrin, *Canadian Evaluation Society Project in Support of Advocacy and Professional Development: Evaluation Benefits, Outputs, and Knowledge Elements* (Oct. 2002), app. C, *Literature Review*, p. 8. <http://consultation.evaluationcanada.ca/englishhome.htm> (Apr. 17, 2009).

Chapter 2: Defining Program Evaluation

This chapter suggests, through a narrative progression, that Supreme Audit Institutions (SAI) may want to adopt a more or less uniform operating concept of program evaluation. First, it notes that SAIs variously define financial audit and performance audit within both the individuality of their disparate national laws and their implementation of the comprehensive framework of INTOSAI's Auditing Standards. Next, the chapter notes that some SAIs include a concept of program evaluation within their definitions of program audit, without expressly defining program evaluation. Then it notes some definitions of evaluation as this term is generally applied by other international programs. Finally, the chapter offers for consideration an operational definition of program evaluation, which, if adopted, may enable SAIs more deliberately to complement their financial and performance audit work with the work of program evaluation.

[2.1] SAIs Variously Define Financial and Performance Audits

Each SAI defines and implements its financial audit function independently, in accordance with national law, but also within the comprehensive framework of INTOSAI's Auditing Standards.¹ Consider only two examples—one from Colombia, one from the People's Republic of China. In Colombia,

[The Political Constitution of 1991, article 267, establishes that “Fiscal control is a public function to be exercised by the Office of the Comptroller

General of the Republic, which oversees the fiscal management of the administration and of individuals or entities which handle funds or assets belonging to the nation.”] ²

According to the National Audit Office of the People’s Republic of China, Article 1 of the Audit Law of the People’s Republic of China— adopted on August 31, 1994, and amended February 28, 2006—states more particularly that

This Law is formulated in accordance with the Constitution, with a view to strengthening State supervision through auditing, maintaining the fiscal and economic order of the country, improving the efficiency in the use of government funds, promoting the building of a clean government and ensuring the sound development of the national economy and society.³

Some SAIs add to their financial audit function the function of performance audit, which they also variously define, largely in accordance with the terms of INTOSAI Audit Standard 1.0.40 on performance audit.⁴ Consider, again, only two examples—one from Australia, one from the Netherlands. The Australian National Audit Office *Performance Auditing in the Australian National Audit Office* states that

A performance audit is an independent, objective and systematic assessment of public sector entities’ programs, resources, information systems, performance measures, monitoring systems and legal and policy compliance.

Performance audits play an important role in improving the administration and management practices of public sector entities.

Performance audits involve the evaluation of the implementation of specific government programs, policies, projects and activities. They also examine how well administrative support systems operate. As such, the audits can include consideration of:

- economy (minimising cost);
- efficiency (maximising the ratio of outputs to inputs);

- effectiveness (the extent to which intended outcomes were achieved); and
- legislative and policy compliance.

There are four types of performance audit:

- audits of a program or activity in a single entity;
- protective security audits (examines security arrangements);
- cross-entity audits (reviews the same activity in a number of entities or the administration of a program by a number of entities); and
- follow-up audits (reviews the implementation of recommendations from a previous audit).

...

Performance audits also identify better practices, which may then be incorporated into Better Practice Guides . . . [and] spread the messages throughout the public sector to help improve the quality of public administration across all entities.⁵

Australia's 2004 *Better Practice in Annual Performance Auditing* adds that

Annual performance monitoring measures are only part of the performance story and generally cannot capture adequately qualitative results, particularly in regard to outcome effectiveness. . . . By undertaking evaluations, and making good use of the findings to explain performance in annual reports, agencies can fill a critical gap in current performance information.

Evaluation addresses questions of whether the “right things” are being done by an agency, as well as whether they are being “done right”. Finance and [the Australian National Audit Office] have encouraged the use of evaluation for almost two decades. Guidance developed by Finance provides a valuable source of advice on how to incorporate evaluation into a performance

management regime.⁶

According to the Netherlands performance audit manual:

Under the [Dutch] constitution, the organisation, composition and powers of the Netherlands Court of Audit are regulated by Act of Parliament, i.e. the Government Accounts Act 2001. The latter Act also contains details of the audits which the Netherlands Court of Audit is entitled to perform. . . .

Under section 85 of the Government Accounts Act 2001, the Netherlands Court of Audit is required to “examine the effectiveness and efficiency of the policy pursued, and the efficiency of financial and material management, of the records kept for this purpose and of the organization of central government.”⁷

The Netherlands manual adds that

[4] INTERNATIONAL EVALUATION COMMUNITY:

the European scene is now an integral part of the international evaluation community. The single European countries may follow their particular paths, but at the same time they are influenced by a new governance model [T]he EU multi-level governance system . . . is a particular mix of both traditions of social programs evaluation and NPM [new public management].

Nicoletta Stame, vice-president of the European Evaluation Society, “Evaluation and the Policy Context: The European Experience,” keynote address, International Conference of the Australasian Evaluation Society, Auckland, New Zealand, September 17, 2003, p. 10. Canadian Evaluation Society, Grey Literature, Ottawa, Ontario. <http://evaluationcanada.ca/site.cgi?s=1> (Apr. 24, 2009).

The efficiency and effectiveness audits conducted by the Netherlands Court of Audit are designed to measure outputs and outcomes. There are a number of different ways of looking at outputs and outcomes, however: we can analyse the degree to which the body in question has achieved its objectives, we can analyse the effectiveness of policy and we can analyse the efficiency of outcomes and outputs.⁸

[2.2] Some Institutions Define Performance Audit Generically to Include Program Evaluation

In July 2007, the Comptroller General of the United States issued a revised edition of *Government Auditing Standards*, often called the “yellow book,” defining financial audits and performance audits:

Performance audits . . . provide assurance or conclusions based on an evaluation of sufficient, appropriate evidence against stated criteria, such as specific requirements, measures, or defined business practices. Performance audits provide objective analysis so that management and those charged with governance and oversight can . . . improve program performance and operations, reduce costs, facilitate decision making by parties with responsibility to oversee or initiate corrective action, and contribute to public accountability. . . . A performance audit is a dynamic process that includes consideration of the applicable standards throughout the course of the audit. . . . Performance audit objectives may vary widely and include assessments of program effectiveness, economy, and efficiency; internal control; compliance; and prospective analyses. These overall objectives are not mutually exclusive.⁹

The 2007 yellow book adds that while auditors commonly follow what are known as generally accepted government auditing standards when they conduct financial or performance audits, for performance audits they

may use other professional standards in conjunction with [these standards],

such as . . . *Guiding Principles for Evaluators*, American Evaluation Association; *The Program Evaluation Standards*, Joint Committee on Standards for Education Evaluation; and *Standards for Educational and Psychological Testing*, American Psychological Association.¹⁰

The National Audit Office of Mauritius, for example, has continually evolved since the early 1960s:

Formerly public auditing and reporting were concerned mainly with two aspects:

- i. Completeness and accuracy of the published statements, and
- ii. Exercise of control.

In the 1960's the Government's financial affairs grew in size and complexity. The improvements in standards of accountancy and the inadequacy of the small establishment of the Audit Office demanded a complete re-appraisal of auditing set up and practices. . . .

Consequently, in quest of professionalism . . . [w]e managed to increase both our technically and professionally qualified personnel

Likewise, our audit approaches have evolved to include programme results reviews/evaluations and since early 1980's we were already in the forefront to implement the modern concept of value for money in our audit work. IT audit and issues relating to Corporate governance have also been taken care by us.

Today we are fully equipped to offer our services, experience and training to other countries in the region¹¹

While Mauritius appears not to use “program evaluation” to describe such activities, at least one report it has issued indicates that this may sometimes generically characterize its work. *Anti - Erosion Programme in Rodrigues* reported on a project aimed at mitigating and controlling soil erosion with the additional objective of improving “the standard of living of the Rodriguan inhabitants by promoting an

equilibrium between the island's economic development and the rational and sustainable use of its natural resources.”¹²

The Mauritius National Audit Office also reported that the Rodrigues review “concentrated on the adequacy of the executing agency in planning, implementing and monitoring of the project rather than on a detailed examination of the financial transactions.”¹³ The evaluators found that although the project was intended to adopt a participatory approach, a strategy for implementing community - level activities was not well - defined, project personnel and technical staff were not sufficiently trained in participatory principles and procedures, and community participation was therefore limited. Moreover, monitoring and evaluation of project activities were lacking.¹⁴ Implicitly, SAI performance audits may thus sometimes be, in some ways, not unlike the evaluations of international development programs.

[2.3] Other International Organizations Define Evaluation

A number of international organizations— whether independent donor institutions, such as The World Bank, or government agencies, such as Danida (Danish International Development Assistance) in the Ministry of Foreign Affairs of Denmark—customarily undertake formal evaluations. In 2003, OECD (Organisation for Economic Co - operation and Development), through its Development Assistance Committee (DAC), recommended a definition of program evaluation that all major donor agencies had adopted:

An evaluation is an assessment, as systematic and objective as possible, of ongoing or completed aid activities, their 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

[5] THE RISE OF PARTICIPATORY EVALUATION:

PPP projects (People's Participation Program by the FAO) have been carried out in 12 countries in Africa, Asia and Latin America. Some 13,200 people have been active PPP participants, while the number of beneficiaries is some 80,000 in total It is important – to the extent possible – for each group to identify, plan, carry out, and evaluate its own activities. . . .“People's Participation Program by the FAO [Food and Agriculture Organization of the United Nations]”, SD dimensions (May 1996).

<http://www.fao.org/sd/spdirect/ppre0006.htm> (April 24, 2009)

both recipients and donors.¹⁵

The definition is implicit in *Evaluation Policy 2006*, a framework intended to communicate Danida's expectations of the quality of evaluations, where noted that

Prior to 1982 evaluations focused essentially on individual projects and programmes. Most of these were mid - term or phase evaluations conducted as the project moved from one phase to the next. Only a few end - evaluations were conducted, and only occasionally were ex - post evaluations carried out to study the long - term effect of projects.

In the period 1982–87, after Danida's Evaluation Unit was established, it was agreed to use evaluations for more systematic studies Also in this period most evaluations were mid - term or phase evaluations of individual projects. The trend was to replace mid - term evaluations with internal reviews and increase the number of end - evaluations. The use of evaluations was more systematic in the sense that it was guided by an annual evaluation programme

During 1987–97 the number of individual project evaluations was reduced and the number of thematic and sector evaluations increased. As a principle, all evaluation reports were made public. In 1992, informing the public was

included as an essential goal for evaluation in accordance with DAC principles. In this period evaluations became more experimental and included a number of impact evaluations as well as use of participatory methods. . . .

A Review of Evaluation in Danida by external international consultants found in 2003 that Danida's evaluation system basically is sound in comparison to general international standards.¹⁶

USAID, the U.S. Agency for International Development, offers on EvalWeb—its Web site dedicated to evaluation practitioners and others interested in evaluation and program assessment—this definition:

An evaluation is . . . a systematic and objective assessment of an on - going or completed project, program, or policy. Evaluations are undertaken to:

- a. improve the performance of existing interventions or policies,
- b. assess their effects and impacts,
- c. inform decisions about future programming

. . .

Evaluation is an analytical effort undertaken to answer specific program management questions and provides a systematic way to gain insights and reach conclusions about the effectiveness of specific activities, validity of a development hypothesis, utility of performance monitoring efforts, factors in the development context that may have an impact on the achievement of results, and the types of actions that need to be taken to improve performance.¹⁷

GAO conducts audits and evaluations under a definition of performance audit, while its standards of accountability in evaluating programs include both retrospective analysis, for effectiveness, efficiency, and economy, and also prospective analysis, as noted in *Government Auditing Standards*:

Prospective analysis audit objectives provide analysis or conclusions about information that is based on assumptions about events that may occur in

the future along with possible actions that the audited entity may take in response to the future events. Examples . . . include providing conclusions based on (a) current and projected trends and future potential impact on government programs and services; (b) program or policy alternatives, including forecasting program outcomes under various assumptions; [and] (c) policy or legislative proposals, including advantages, disadvantages, and analysis of stakeholder views¹⁸

[2.4] An Operational Definition of Program Evaluation for SAIs

Addressing issues of definition, as well as practice, the Canadian Evaluation Society (CES) in October 2002 published a report of a project in which the society had attempted to develop a core body of knowledge for program evaluation. CES had attempted to identify program evaluation's benefits and the knowledge and skills needed to produce them.

Among the society's summary "considerations" were that "there is no universally accepted definition of program evaluation" but that

some basic characteristics of evaluation . . . appear to be widely accepted:

- Evaluation applies research design principles to answer practical questions about programs.
- Data is collected and processed systematically to provide evidence about what is happening in a program (processes and outcomes), why it is happening, and how the program can be improved. This evidence can be used to make judgements about the program's merit or worth.
- Evaluation is cross-disciplinary and draws methods from many different fields of study.¹⁹

In CES's literature review, the society reprinted several definitions of program evaluation, all characterized by their emphasis on improving program effectiveness in order to benefit society. Here are three of those definitions.

First, a definition by William Shadish, Thomas Cook, and Laura Leviton from their 1995 *Foundations of Program Evaluation*:

Intrinsic to evaluation is an idealized problem – solving sequence for (a) identifying a problem; (b) generating alternatives to reduce its symptoms; (c) evaluating these alternatives; and then (d) adopting those that results suggest will reduce the problem satisfactorily.²⁰

Second, a definition by Eleanor Chelimsky and William Shadish from their 1997 *Evaluation for the 21st Century*, speaking of three key areas of benefits:

These different purposes, along with the questions they seek to address, seem to fall naturally into three general perspectives:

- Evaluation for accountability (e.g., the measurement of results or efficiency)
- Evaluation for development (e.g., the provision of evaluative help to strengthen institutions)
- Evaluation for knowledge (e.g., the acquisition of a more profound understanding in some specific area or field).²¹

Third, a definition by Peter Rossi, Howard Freeman, and Mark Lipsey from their 1999 *Evaluation: A Systematic Approach*:

evaluation researchers use social research methods to study, appraise and help improve social programs in their important aspects, including the diagnosis of the social problems they address, their conceptualization and design, their implementation and administration, their outcomes and their

efficiency.²²

Consider finally, in this context, how program evaluation evolved at the U.S. Government Accountability Office, as indicated in a definition from 1992 and another from 2005. In 1992, GAO defined program evaluations in *The Evaluation Synthesis* as “The application of scientific research methods to assess program concepts, implementation, and effectiveness.”²³ Thirteen years later, in *Performance Measurement and Evaluation: Definitions and Relationships*, GAO redefined program evaluations as

individual systematic studies conducted periodically or on an ad hoc basis to assess how well a program is working. They are often conducted by experts external to the program, either inside or outside the agency, as well as by program managers.

A program evaluation typically examines achievement of program objectives in the context of other aspects of program performance or in the context in which it occurs. Four main types can be identified, all of which use measures of program performance, along with other information, to learn the benefits of a program or how to improve it.²⁴

The document goes on to define the four main types as process (or implementation) evaluation, outcome evaluation, impact evaluation, and cost - benefit and cost - effectiveness analyses.²⁵

Taking all this into consideration, SAIs might adopt the following as a general definition of program evaluation:

A program evaluation is a systematic investigation of an organization or institution, program or project, or process or policy that is intended to benefit society. A program evaluation’s purpose may be to help improve, help decide an action, or learn reasons for successes and failures or strengths and weaknesses in meeting objectives. It can document for accountability or increase knowledge. Program evaluations commonly determine criteria that distinguish between high and low quality, measure performance against those

criteria, and draw valid and reliable evaluative conclusions.

A program evaluation may be independent while being conducted with participation from community members, consumers, managers, and others to give it multiple perspectives. The evaluation process may involve identifying objectives and key questions, defining assessment criteria, obtaining and analyzing relevant data, reporting results, and promoting the effective use of the findings. Evaluations are conducted by persons well-grounded in research design and methodology, content knowledge and skills in particular specializations, and competencies in evaluation theory, evaluation methodology, and practical knowledge.²⁶

-
- 1 See Inga - Britt Ahlenius, Auditor General, Swedish National Audit Office, “Auditing Standards and Ethics: An INTOSAI Perspective,” address, 17th Annual Conference on New Developments in Government Financial Management for Government Financial Managers, Miami, Florida, April 2003. International Consortium on Governmental Financial Management. www.icgfm.org/downloads/presentations/IngaBritt_Ahlenius.ppt (Apr. 18, 2009).
 - 2 Colombia, Contraloría General de la República, Institucional, ¿Qué es la Contraloría? www.contraloriagen.gov.co/html/home/home.asp (Apr. 18, 2009). English translation, University of Richmond, Constitution Finder, Constitution of Colombia (1991), title X, ch. 1, article 267. <http://confinder.richmond.edu/index.php> (Apr. 18, 2009).
 - 3 People’s Republic of China, National Audit Office, Laws and Regulations, Audit Law, Audit Law of the People’s Republic of China, p. 2. www.cnao.gov.cn (Apr. 18, 2009).
 - 4 See INTOSAI Auditing Standards Committee, *Auditing Standards*, 14th INTOSAI Congress, Washington, D.C., 1992; restructured by the 17th INTOSAI Congress, Seoul, Korea, 2001, standard 1.0.40 on performance audit, ch. 1, p. 33. INTOSAI: International Organization of Supreme Audit Institutions; Committees, Working Groups, Task Forces; Goal 1 Professional Standards; Documents Audit - related; Code of Ethics and Auditing Standards. www.intosai.org (Apr. 18, 2009).
 - 5 Australian National Audit Office, *Performance Auditing in the Australian National Audit*

- Office* (Canberra: 2008), p. 3. Australia National Audit Office, Publications, Other Publications. www.anao.gov.au/index.cfm (Apr. 26, 2009).
- 6 Australia National Audit Office, Publications, *Better Practice Guides, Better Practice Guide: Better Practice in Annual Performance Reporting* (Canberra: April 2004), p. 37. www.anao.gov.au/index.cfm (Apr. 26, 2009).
 - 7 Netherlands Court of Audit, *Performance Audit* (The Hague: Feb. 2, 2006), ch. 2, sec. 2.2.1, p. 14. Netherlands Court of Audit, Publications, Manuals. www.rekenkamer.nl/9282400/v (Apr. 20, 2009). A footnote, adding “The legislation does not actually make clear what exactly is meant by ‘the efficiency of financial and material management, of the records kept for this purpose and of the organisation of central government,’” discusses this.
 - 8 Netherlands Court of Audit, *Performance Audit*, ch. 1, sec 1.1, p. 8. The only definition of performance audit appears in the manual’s definition of performance data (annex 9, p. 197, emphasis in original): “There are three types of performance data: policy objectives, which should be operationalised as far as possible in terms of desired and actual outcomes, linked to certain target groups and to a specified time horizon; information on the cost price and quality of *outputs* delivered or to be delivered by the government (i.e. products and services); [and] programme expenditure, supported as much as possible by *volume and price data*”
 - 9 Comptroller General of the United States, *Government Auditing Standards, July 2007 Revision*, GAO - 07 - 731G (Washington, D.C.: U.S. Government Accountability Office, July 2007), secs. 1.25–1.28, pp. 17–18. www.gao.gov/govaud/ybk01.htm (Apr. 20, 2009).
 - 10 Comptroller General of the United States, *Government Auditing Standards, July 2007 Revision*, sec. 1.03, p. 12.
 - 11 Mauritius, National Audit Office, The Office, History, The Development of Audit. www.gov.mu/portal/site/auditsite (Apr. 20, 2009).
 - 12 Mauritius, National Audit Office, Reports and Publications, *Annual Report 1999–2000*, vol. 2, *Special Reviews: Ministry for Rodrigues*. Rodrigues Administration, Anti - Erosion Programme in Rodrigues (Port Louis: 2000), sec. 4.1.4.

- www.gov.mu/portal/site/auditsite (Apr. 26, 2009).
- 13 Mauritius, National Audit Office . . . *Ministry for Rodrigues*. Rodrigues Administration, *Anti Erosion Programme in Rodrigues*, sec. 4.1.1.
 - 14 Mauritius, National Audit Office . . . *Ministry for Rodrigues*. Rodrigues Administration, *Anti Erosion Programme in Rodrigues*, executive summary.
 - 15 Hans E. Lundgren, W. Haven North, and Ray C. Rist, *A Review of Evaluation in Danida Ministry of Foreign Affairs* (Copenhagen: January 2003), p. 8, fn 1. See Ministry of Foreign Affairs of Denmark, Development Policy, Evaluation of Development, About the Evaluation Department. www.um.dk/en (Apr. 24, 2009).
 - 16 Ministry of Foreign Affairs of Denmark, Development Policy, Evaluation of Development, Evaluation Policy & Guidelines, *Evaluation Policy 2006* (Copenhagen: Evaluation Dept., 2006), p. 1. www.um.dk/en (Apr. 24, 2009).
 - 17 USAID EvalWeb, What is an evaluation? <http://evalweb.usaid.gov/> (Apr. 24, 2009).
 - 18 Comptroller General of the United States, *Government Auditing Standards*, sec. 1.32, p. 22. See also GAO, *Prospective Evaluation Methods: The Prospective Evaluation Synthesis*, PEMD - 10.1.10 (Washington, D.C.: Nov. 1990). www.gao.gov/special.pubs/pe10110.pdf (Apr. 24, 2009).
 - 19 Rochelle Zorzi, Martha McGuire, and Burt Perrin, *Canadian Evaluation Society Project in Support of Advocacy and Professional Development: Executive Summary* (Toronto: Oct. 2002), p. ii. <http://consultation.evaluationcanada.ca/english/home.htm> (Apr. 24, 2009).
 - 20 William R. Shadish Jr., Thomas D. Cook, and Laura C. Leviton, *Foundations of Program Evaluation: Theories of Practice* (Thousand Oaks, Calif.: Sage Publications, 1995), p. 21, in Zorzi, McGuire, and Perrin, *Canadian Evaluation Society Project in Support of Advocacy and Professional Development*, app. C, p. 2.
 - 21 Eleanor Chelimsky and William Shadish, *Evaluation for the 21st Century: A Handbook* (Thousand Oaks, Calif.: Sage Publications, 1997), p. 10, in Zorzi, McGuire, and Perrin, *Canadian Evaluation Society Project in Support of Advocacy and Professional*

Development, app. C, p. 2.

- 22 Peter H. Rossi, Howard E. Freeman, and Mark Lipsey, *Evaluation: A Systematic Approach* (Thousand Oaks: Calif.: Sage Publications, 1999), p. 21, in Zorzi, McGuire, and Perrin, *Canadian Evaluation Society Project in Support of Advocacy and Professional Development*, app. C, p. 2.
- 23 GAO, *The Evaluation Synthesis*, PEMD – 10.1.2 (Washington, D.C.: Mar. 1, 1992), p. 126. www.gao.gov (Apr. 24, 2009).
- 24 GAO, *Performance Measurement and Evaluation: Definitions and Relationships*, GAO - 05 - 739SP (Washington, D.C.: May 2005), p. 2. www.gao.gov (Apr. 24, 2009).
- 25 GAO, *Performance Measurement and Evaluation: Definitions and Relationships*, p. 4. Types of evaluation are discussed below, in chapter 4.
- 26 Adapted from Western Michigan University, The Evaluation Center, About Us, About the Center, Kalamazoo. www.wmich.edu/evalctr/about.html (Apr. 24, 2009).

Chapter 3: Planning for Program Evaluation

Program evaluations are planned by understanding why an evaluation's requesters want to have the evaluation performed, what they intend to do with the results, and who the audience will be. Planning is also a matter of gaining insight into a program

[6] ON INTERACTIVE COMMUNICATION:

Auditing carried out by independent national audit offices is an institutionalised part of the policy cycle of most governments in the developed world. It involves the evaluation of financial accounts, financial management and— increasingly—of results of public policies and programs. . . . The independent position of most national audit institutions also implies relatively separated worlds of auditors on one hand and auditees on the other. . . . [I]ntensive communication between auditors and auditees in different stages of the audit process seems an important success factor.

Gerard Bukkems and Hans de Groot, Netherlands Court of Audit, "Learning by Auditing: Mission Impossible?" Paper for the 5th biennial conference of the European Evaluation Society, Three Movements in Contemporary Evaluation: Learning, Theory and Evidence, Seville, October 10-12, 2002, p. 5.

www.evaluationcanada.ca/distribution/20021010_bukkems_gerard_de_groot_hans.pdf (Apr. 27, 2009).

evaluation's purpose and objectives, formulating questions about the program, and making sure that the evaluation team has the education, skills, and experience necessary to complete the evaluation's tasks.

The process is interactive. It is accomplished by the evaluating organization, the evaluation team, the program managers, others who have a stake in the evaluation's outcome, and sometimes the media talking together about why and how evaluation should be conducted. Therefore, program evaluation differs from financial audit and performance audit; it entails a different way of thinking, working, and communicating, although the process of auditors and auditees communicating also has its acknowledged benefits.¹

It has also been said that "The principal challenge" of program evaluation, regardless of the evaluators' belief in a program's educational or social worth, is "the need to make available that programme's experience for public learning."² Those who have a stake in evaluations—their stakeholders—include people outside the evaluation team who contribute expertise to an evaluation project as well as people the evaluation affects, among whom are participants in the programs being evaluated.

Financial audits primarily raise questions about whether and what data exist, while in program evaluation, the evaluating organization's staff plan an evaluation, which includes framing a design. A design's elements are generally the kinds of information evaluators collect about a program and how and where they collect it, the methods they use for sampling the data, their basis for making comparisons, and the plan they make for analyzing the data. Naturally, the evaluation team's arriving at an appropriate design depends on the content knowledge, competence, and skills of the staff who make up the team.

This chapter illustrates briefly some of the contrasts between financial and performance audit and program evaluation from a planning perspective. It shows program evaluation planning as an evaluation team's using its skills and experience collaboratively with the evaluation's requesters to formulate evaluation objectives. It also shows that the evaluation team chooses a design strategy ultimately to expand what is known about a policy, agency, or program and the people who participate in

or benefit from it.

[3.1] Working with Evaluation's Stakeholders

Organizations that traditionally conduct financial audits or performance audits and that expect to plan program evaluations may look toward a new perspective and begin to build a new set of skills. Unlike the literature on program evaluation, mandates and guides for planning financial audits and performance audits seldom mention the need for collaborating with the people they audit or the people who use the results of their audit work.

Not atypical, for example, is the Office of the State Audit of Vietnam in requiring audit teams who conduct financial, compliance, and performance audits “to implement audit of the right audited entity and in line with the objectives, contents, scope, location and time limit of the audit as set forth in the audit decision by the State Auditor General” and to

use audit professional methods to collect and judge audit evidences; to do checks, comparisons and certifications; to investigate organisations and individuals relating to the audit activities to set foundations for forming comments, certifications, conclusions and recommendations on the contents audited.³

Mention in the law is made of collaborating with the auditees and the audit's users neither here nor in a chapter of the state audit law on audits once they are completed:

The audited entities shall implement the audit conclusions and recommendations . . . on irregularities in the financial statements and mistakes concerning the compliance with law; implement corrective methods over weaknesses in their activities detected and recommended by the State Audit.

Agencies or competent persons who use the audit conclusions shall decide

themselves on the acceptance of audit conclusions of the State Audit and be responsible to law for such decisions.

Once accepted by agencies and/or competent persons, the audit conclusions shall be obliged to follow.⁴

[7] A NORM FOR AUDIT PLANS:

The Performance Audit Manual of the Supreme Audit Institution of Bangladesh—chapter 3 on “Planning Individual Performance Audit”—prescribes that evaluators complete the following checklist as they plan their approach to individual audits:

- Background information on the proposed audit area
- Main policy objectives and instruments for implementation (can be obtained from various policy documents and strategic plans)
- Any Parliamentary interest or previous audit involvement
- An assessment of the risks to achieving VFM [value for money]
- Proposed scope of the study and main issues to be addressed
- Whether there are any other examinations proposed in the same area or which might overlap
- The proposed methodology (the approach and techniques to be used) and sources of evidence
- Expected impacts of the study in terms of savings or improvements
- Staffing requirements and cost to carry out the study
- The proposed timing of the study.

Comptroller and Auditor General of Bangladesh, Performance Audit Manual (Dhaka, Bangladesh: Dec. 26, 2000), p. 25. Office of the Comptroller and Auditor General of Bangladesh, www.cagbd.org/html/perfaudit.html (Apr. 27, 2009).

Not atypical, however, is reform away from audit courts' primarily exercising

1—Control over the accounts, which are submitted by public accountants [aimed] at controlling the legality of collection operations and their conformity to collection laws, and of the legality of payment documents and accounts.

2—Control over financial officers, who manage public funds [whose] objective is to penalize the violators of financial laws and regulations.⁵

and toward focusing on the auditees' improvement through activities

more varied and broad in scope as to encompass Performance Audit. Such an Audit ensures that the auditee's resources management is conducted in an efficient manner and in compliance with rational economic standards and methodologies. Furthermore, it helps monitor how much success has been realized in the application of set working plans and hoped for objectives.⁶

Reference to limited consultation with auditees may be found in several SAIs' manuals. *Australia's General Guidance on the Conduct of Performance Audits*, for example, states that an audit's opening interview

enable[s] the audit team to meet key staff of the audited body; [and]

- further explain the type of audit to be conducted and the objectives, scope, timing and methods to be employed;
- provide an opportunity for the audited body to comment/make suggestions for ANAO [Australian National Audit Office] consideration;
- establish suitable liaison arrangements at both the *management and working levels*, including arrangements for progressive advice and discussion of tentative findings;
- ensure the audited body clearly understands the audit processes . . .⁷

Australia's guidance adds that "It is particularly important that senior program managers are involved in . . . discussions at all stages of the audit to ensure that the audit team is fully aware of the environment surrounding, and issues affecting, the function under review."⁸

In New Zealand, in the 2002 edition of the Auditor - General's performance auditing manual, movement toward broader practices and conventions was initiated in a section called "As performance and accountability become more complex, external accountability reporting needs to change":

One view of accountability is that the content of a report should be confined to matters which are within the control of the preparers and for which they are directly accountable. However, this interpretation of accountability can deprive stakeholders of important information.

...

The best external accountability reporting draws selectively from a comprehensive set of performance elements to identify, measure and report those performance elements of interest to stakeholders. Compliance with statutory performance reporting requirements is essential, but statutes do not limit or prohibit enhanced reporting.

Reflecting a comprehensive model in external accountability reporting also requires a range of specific capabilities. Organisational capability can be built in respect of: understanding of operations and of stakeholder information requirements, in order to identify which elements of performance should be reported and which performance indicators or measures will be relevant to report on those elements⁹

The stated rationale for enhancements to performance audit was "to contribute to improvement in public entities' accountability" and to indicate that "A framework for external accountability reporting must take all elements of performance into

consideration for potential reporting; and report performance elements that stakeholders want to know.”¹⁰

Drawing participants into evaluation can be seen, for instance, in the 2002 report, *The town policy*, from the Cour des Comptes of France that partially crossed over between performance audit and program evaluation. The evaluators first consulted with program participants—senior civil servants, elected local managers, and members of nonprofit organizations—to design the evaluation of a national urban renewal program. Then they visited seven sites in four regions of the country to meet with elected councilors, officials of state services, and members of associations who were also program participants.¹¹ Thus, stakeholders included government officials, nongovernmental organizations, and civil society at large.

A 2003 paper noted the importance of others in planning evaluations. In *Evaluation: A Trade Union Perspective*, Bandula Kothalawala explained that “in trade union circles, evaluation is . . . incorporated into the project document at the design stage and forms an essential element of it.”¹²

After noting that the objectives of trade union evaluations are predominantly about the development of human resources, he added:

Efforts are made to involve as many stakeholders as possible in the [evaluation] process, with priority being given to primary beneficiaries. Collaborative nature of exercise is stressed and perceptions of prime beneficiaries and other stakeholders are given adequate weight when analysing data. It is customary to organise evaluation workshops to be attended by stakeholders. These are normally held in collaboration with local partner organisations in locations convenient for stakeholders. All participants are advised of details of programmes and arrangements are made well in advance. Workshop participants review the achievements of the project in relation to its objectives and add a qualitative and subjective dimension to the whole process, often complementing information obtained through other methods.

It is also common to conduct semi - structured informal interviews with a

representative sample of prime beneficiaries during an evaluation mission.¹³

In a footnote, he described some of the evaluators' interactions during data collection:

Interviews with health professionals during visits to local clinics and hospitals in Western Cape, South Africa in the course of the evaluation of the Training of Health Professionals in the Diagnosis of Occupationally related Diseases provided a good opportunity to the evaluator to acquaint himself with the situation at the local level.¹⁴

Last are two examples from program evaluators speaking about their work while in Honolulu, Hawaii, at the American Evaluation Association's annual meeting in 2000. At session 111, Consumers as Researchers: Tools for Incorporating Mental Health/Substance Abuse Treatment Consumers into the Evaluation Process, "Evaluators and consumers on a federal Substance Abuse, Mental Health, Traumatic Violence research project [led] a discussion on strategies for effectively incorporating consumers (and anyone new to research) into the research team."¹⁵

[8] EVALUATION AND THE STAKEHOLDERS:

For the Walloon Evaluation Society (Société Wallonne de l'Evaluation et de la Prospective), evaluation consists of the analysis — in terms of relevance, efficacy, efficiency, consistency, sustainability and viability – of the policies, the programs, the projects or the functions, conducted with a view to improving the quality of decisions, to better allocating resources and to accountability to the citizens. Evaluation is designed as a participatory and democratic approach involving in the evaluation process any person or group that feels concerned by evaluation.

Société Wallonne de l'Evaluation et de la Prospective, Documentation, Evaluation. www.la-swep.be/index.php (Apr. 27, 2009).

Reasons given for this approach were that “Consumers have a unique perspective that can help guide all aspects of the evaluation process.” The discussion included a brief description of a culturally specific creativity project and elements of consumer involvement in the project, “including: benefits to research, benefits to consumers, pathways and barriers to effective consumer involvement, as well as potential solutions to those barriers.”¹⁶

Similarly, during session 115, at a roundtable presentation entitled “Clinical and Political Implications in Evaluating a Bicultural Program,” Gerald A. Bunn began by describing the evaluation’s context:

It is necessary, when evaluating Alcohol & Drug programmes with a cultural component, to pay attention to political as well as clinical issues. Although many ethnic groups now live in New Zealand, the main cultural division is still identified as Maori (New Zealand's “tangata whenua”, or “people of the land”) and Pakeha (non – Maori). The status of Maori is defined in the Treaty of Waitangi, considered the nation's founding document. The Treaty is a unique political instrument which aims to protect and enhance “rangatiratanga” (sovereignty), which is the right of Maori to live and develop in a Maori way. This has been interpreted as meaning that Maori should have control over their own health services. Generally, New Zealand’s mental health services are delivered as either Maori – only or mainstream programmes.¹⁷

Accordingly, he said, the hospital alcohol and drug treatment program being evaluated was a “parallel program,” “whereby Maori are treated in their own cultural setting, the Taha Maori programme,” and “Pakeha, or non – Maori, are treated using a mainstream approach,” so that the two programs were evaluated separately. However, it had been

recognised that “instruments developed along Eurocentric approaches cannot be blindly applied to people of colour.” . . . [and that] the absence of specific instruments and procedures for Maori required adaptation of Western tools and procedures. To accomplish this, extensive and time consuming consultation was required with a variety of stakeholders.¹⁸

[9] ON LEARNING FROM PROGRAM EVALUATION:

The five agencies we reviewed employed various strategies to obtain useful evaluations of program effectiveness. Just as the programs differed from one another, so did the look and content of the evaluations and so did the types of challenges faced by agencies. . . . Whether evaluation activities were an intrinsic part of the agency's history or a response to new external forces, learning from evaluation allowed for continuous improvements in operations and programs, and the advancement of a knowledge base. . . .

GAO, Program Evaluation: An Evaluation Culture and Collaborative Partnerships Help Build Agency Capacity, GAO-03-454 (Washington, D.C.: May 2, 2003), p. 24. www.gao.gov (Apr. 27, 2009).

In other words, in planning evaluations, program evaluation teams generally try to discuss with participants and others why the evaluation is being done and its objectives; the characteristics of the program; the preliminary evaluation questions; any assumptions about the program's problems and possible solutions; and any evaluation alternatives.

According to the U.S. Government Accountability Office, for example,

reaching agreement with the sponsors, users, program operators, and others on the contents and implications of a question can be difficult and challenging. Among the several reasons for the strenuousness of the task is that the formulation of a problem has preeminent importance in the remaining phases of the evaluation.¹⁹

In brief, the context of financial and performance audit is, in general, accountability and compliance, while program evaluation's context extends to interactions that produce lessons learned from what may be complex and cross-cutting policy issues.

[3.2] Adopting a Design Strategy

In adopting a design strategy, skilled evaluators understand that it is not necessarily the particular method they choose that defines the strength of their evaluation:

Longitudinal, experimental, quasi – experimental, before – and – after, and case study evaluations can be either strong or weak. . . . A simple before – and – after design without controls will always present problems of internal validity. Yet sample surveys and control groups can be impossible for a variety of reasons. That is, the strength of an evaluation has to be judged within the context of the question, the time and cost constraints, the design, the technical adequacy of the data collection and analysis, and the presentation of the findings.²⁰

“Designing an Evaluation,” chapter 3 of the European Commission’s 2004 *Evaluating EU Activities: A Practical Guide for the Commission Services*, details, in both text and charts, how SAIs may plan for and design program evaluations. The guide describes drafting an evaluation project’s context and objectives as well as its timetable and management, plus drawing up evaluation questions and terms of reference—a document that describes the activity being evaluated, specifies the information available to the evaluators, states a number of questions the evaluation has to answer, sets out the steps of the evaluation process, and outlines the evaluation method.²¹

With respect to design strategy, the Commission’s guide reminds SAIs that

many different evaluation methods [are] possible Hence, in the case of an external evaluation, it can be in the interest of the commissioner of the evaluation to provide only a general outline, and thus invite the potential evaluator to propose an appropriate solution, depending on his/her experience

and expertise. . . .

Other factors also influence the design of the evaluation, notably the time and resources available for the evaluation project. Bigger budgets and longer time horizons enable the commissioner of the evaluation to ask for more varied and in - depth data to be collected and for more sophisticated analyses to be made. The possibilities and restrictions that follow from the size of the budget can be summarised as follows:

- Short time span and small budget: desk studies, interviews, focus groups.
- Medium time span and budget: case studies, surveys, expert panels.

- Long time span and considerable budget: econometric models, cost - benefit and cost effectiveness analyses.²²

Before actually executing a program evaluation, the evaluation team considers the kinds of information it will collect and its sources, the sampling methods, the basis for making comparisons with and without a program, and the analysis plan. These and other choices the evaluators make for the design's construction will determine the quality of the information they can obtain, the strength of the conclusions they can draw, and the evaluation's cost, timeliness, and usefulness.

In short, the evaluation team within an organization that plans program evaluations strategizes each evaluation project's methodology. The team writes an evaluation

[10] EVALUATION DESIGN AND PROGRAM EFFECTS:

Finally [besides the design constraints of budget and time], other aspects of an evaluation have important repercussions on the choice of the evaluation method, such as:

- Does the evaluation concern a programme or an activity addressing clearly identified beneficiaries or is it a policy that addresses more or less indirectly a broad and potentially difficult-to-contact public?
- Will the evaluation focus just on the direct, shorter-term effects on addressees (e.g. as in the case of a mid-term evaluation) or will more indirect effects, longer-term effects and unintended effects be addressed?

European Commission, Directorate- General for the Budget, Evaluating EU Activities: A Practical Guide for the Commission Services (Luxembourg: Office for Official Publications of the European Communities, July 2004), p. 45. European Commission, Financial Programming and Budgeting, Publications. http://ec.europa.eu/budget/index_en.htm (Apr. 27, 2009).

plan that is intended to produce results of the best possible quality with respect to the program and also to help develop the organization's future evaluative capacity.

[3.3] Linking Design to Evaluation Questions

Whether formulated in terms of policy goals or program goals, an evaluation's objectives are best framed as questions that all have agreed, collaboratively with the evaluators, that the evaluators should work to answer. A program evaluation's questions are generally descriptive or normative or causal. In this respect, they emphasize areas rather different from areas that audit questions emphasize.

Moreover, an evaluation design evolves from evaluators' linking evaluation questions to what appears to be the most appropriate design choice. They do this by focusing each evaluation question ever more narrowly as they make a series of design - screening decisions. Beginning with the evaluation question, the evaluation team will ask whether it is

- descriptive (perhaps it asks about how a training program was implemented) or
- normative (asking whether the program's job – placement goals have been met) or
- causal (asking how the program has affected job – placement rates).

Evaluators focus the design still more clearly by looking at how definitive stakeholders expect the conclusions to be and at the kinds of constraints they expect the project to have. For example, they might clarify whether or not generalizations will be made from what they find to some larger population or class of things. Now they can link the evaluation question to a series of design questions, like

- Can a comparison group be formed?
- Can subjects be randomly assigned to groups?

- Can outcomes be measured over time?

And, finally, How much time is available to conduct the evaluation?

With this type of process, a design gradually evolves. In arriving at a final design, evaluators balance technical considerations against constraints, as the evaluation questions become more specific and the research possibilities become narrower:

For example, it might be necessary to choose between collecting new data, which might answer the evaluation questions comprehensively, and using available data, which is usually the least expensive course and the quickest but may leave some avenues unexplored.²³

The kinds of objectives that are posed in the evaluation questions and that evaluators focus and scope into a methodological framework for study and analysis may relate to any number of areas: how government can improve information, the need for a new or changed policy, a program's policy context, or a program's net effect, among other things. Some specific questions that evaluators might thus be faced with in creating an evaluation design could be

- Are government evaluations technically sound? Are the performance indicators meaningful? What new strategies could be used to assess policy?
- What key aspects of the current context affect program policy issues? Which stakeholder views are relevant to policy and program issues?
- Is an impact assessment needed? What are the future needs for a new law or regulation?
- What effect is the program having beyond what would have happened if the program did not exist? What is the relationship between the program's costs and its estimated benefits?²⁴

Such questions differ from questions that tend to be asked in financial or

performance audits:

- Are information systems complete, accurate, consistent, and without fraud?
- Are program operations using their resources efficiently to deliver services?
- Is the program achieving its stated goals and objectives?
- Has the program been implemented according to law, policy, and procedures?

[3.4] Assessing a Team’s Capacity to Execute Evaluation Plans

The W. K. Kellogg Foundation’s *Evaluation Handbook* indicates that

the primary goals of evaluation are that stakeholders are engaged, active participants . . . and that the evaluation process and findings will be meaningful and useful to those ultimately responsible for improving and assessing the program. In the end, this means that there is no one way to do evaluation.²⁵

The *Evaluation Handbook* suggests further that, depending on the relationship between primary stakeholders and the evaluation team,

If the evaluation purpose is to determine the worth or merit of a program, you might look for an evaluator with methodological expertise and experience. If the evaluation is focused on facilitating program improvements, you might look for someone who has a good understanding of the program and is reflective. If the primary goal of the evaluation is to design new programs based on what works, an effective evaluator would need to be a strong team

player with analytical skills.

Experience tells us, however, that *the most important overall characteristics to look for in an evaluator are the ability to remain flexible and to problem – solve.*²⁶

Institutions that conduct program evaluations may assess their probabilities for success by looking at the context of the programs they evaluate. In selecting evaluation designs and methodologies, they effectively determine the knowledge, skills, and abilities they require from their evaluation teams. Before moving on to execute an evaluation study, SAIs are wise, therefore, to take a look, with their evaluation teams, at each design, as well as the overall evaluation plan, to gain total

[11] EVALUATORS' KNOWLEDGE, SKILLS, AND ABILITIES:

Evaluators should possess or ensure the provision of content knowledge appropriate for the evaluation and continuously strive to improve their methodological and practice skills. Evaluators should possess the knowledge, skills and experience in:

- the application of sound research design able to answer the chosen questions;
- the collection and analysis of reliable quantitative and qualitative data; and
- the development of valid, credible and unbiased conclusions and recommendations.

Treasury Board of Canada Secretariat, Centre of Excellence for Evaluation, Evaluation Function in the Government of Canada (Ottawa: July 6, 2004), appendix 2, Evaluation Standards in the Government of Canada. Treasury Board of Canada Secretariat, Organization, Evaluation, Tools and Resources, Historical Documentation. www.tbs-sct.gc.ca/tbs-sct/index-eng.asp

perspective on a study's implications, not only for those who benefit from its findings but also for the evaluating institution. Among these implications are issues like these:

- what design elements the evaluators have chosen and left out;
- whether they have selected a design that will appropriately answer the study questions;
- whether they have made an evaluation plan to address the study's constraints;
- how the evaluators intend to account for the study's strengths and weaknesses;
- stakeholder expectations and the evaluators' expectations for meeting them;
- how the evaluation's lessons will help the evaluators develop new skills and capacities for themselves, their organization, and the evaluation community.

Resources and ideas for responding to such issues are outlined in chapter 5.

- 1 See Gerard Bukkems and Hans de Groot, Netherlands Court of Audit, "Learning by Auditing: Mission Impossible?" Paper for the 5th biennial conference of the European Evaluation Society, Learning, Theory and Evidence, Seville, October 10– 12, 2002, p. 5. www.evaluationcanada.ca/distribution/20021010_bukkems_gerard_de_groot_hans.pdf (Apr. 27, 2009).
- 2 Saville Kushner, "The Object of One's Passion: Engagement and Community in Democratic Evaluation," Keynote address, Australasian Evaluation Society International Conference, Wollongong, Australia, October–November 2002, p. 8. Australasian Evaluation Society, Conferences, Previous. www.aes.asn.au (Apr. 27, 2009).
- 3 Vietnam, Office of the State Audit, Socialist Republic of Vietnam, National Assembly of

- the Socialist Republic of Vietnam, State Audit Law (Dec. 25, 2001), ch. 4, sec. 5, article 52. www.kiemtoannn.gov.vn/ktnnEnglish/thongtinM.htm (Apr. 27, 2009).
- 4 Vietnam, Office of the State Audit, Socialist Republic of Vietnam, National Assembly of the Socialist Republic of Vietnam, State Audit Law (Dec. 25, 2001), ch. 1, article 9.
 - 5 Lebanese Republic, Court of Audit, About COA, Goals and Mission, Jurisdiction and Procedures of the Audit Court. www.coa.gov.lb (Apr. 27, 2009).
 - 6 Saudi Arabia, General Auditing Bureau, The Kingdom of Saudi Arabia's Message to the 7th Global Forum on Building Trust in Government for Further Participation and Transparency, "The Saudi's Message in Brief," p. 12. www.gab.gov.sa/index.php (Apr. 27, 2009).
 - 7 Australia National Audit Office, *General Guidance on the Conduct of Performance Audits* (Canberra: July 2003), sec. 25, p. 4. Australia National Audit Office, Publications, Other Publications. www.anao.gov.au/ (Apr. 27, 2009).
 - 8 Australia National Audit Office, *General Guidance on the Conduct of Performance Audits*, sec. 25, p. 5.
 - 9 New Zealand, Report of the Controller and Auditor - General, *Reporting Public Sector Performance*, 2nd ed. (Wellington: Jan. 2002), pp. 22–23. New Zealand Office of the Controller and Auditor - General, Good Practice Guides. www.oag.govt.nz (Apr. 27, 2009). Emphasis in original.
 - 10 New Zealand, Report of the Controller and Auditor - General, *Reporting Public Sector Performance*, p. 5.
 - 11 France, Cour des Comptes, *La Politique de la Ville* (Paris: Feb. 2002), see esp. pp. 12–14 and 206–07. France, Cour des Comptes, Les juridictions financiers, Publications, Rapports publics thématiques. www.ccomptes.fr/ (Apr. 27, 2009).
 - 12 Bandula Kothalawala, *Evaluation: A Trade Union Perspective* (London, Eng.: Trades Union Congress, European Union and International Relations Department, n.d.), p. 4. www.minefi.gouv.fr/fonds_documentaire/TRESOR/cicid/atelier/contrib/27.pdf. (Apr. 27, 2009).

- 13 Kothalawala, *Evaluation*, p. 5.
- 14 Kothalawala, *Evaluation*, p. 6.
- 15 See session 111 (Wed. 4:30), Consumers as Researchers: Tools for Incorporating Mental Health/Substance Abuse Treatment Consumers into the Evaluation Process, Chaired by Debi Elliott, Portland State University. American Evaluation Association, Training, Conference History, Evaluation 2000, November 1–4, Honolulu. www.eval.org (Apr. 27, 2009).
- 16 Session 111, Consumers as Researchers, Chaired by Debi Elliott. American Evaluation Association, Training, Conference History, Evaluation 2000.
- 17 See session 115 (Wed. 4:30), Examining Treatment and Prevention Programs, Robert G. Orwin, host, Battelle Centers for Public Health Research Considering Culture in Evaluating Health Programs, Gerald A. Bunn, “Clinical and Political Implications in Evaluating a Bicultural Program.” American Evaluation Association, Training, Conference History, Evaluation 2000, November 1–4, Honolulu. www.eval.org (Apr. 27, 2009).
- 18 Session 115, Examining Treatment and Prevention Programs, Gerald A. Bunn, “Clinical and Political Implications in Evaluating a Bicultural Program.” American Evaluation Association, Training, Conference History, Evaluation 2000.
- 19 GAO, *Designing Evaluations*, PEMD - 10.1.4 (Washington, D.C.: May 1991), p. 11. www.gao.gov (Apr. 27, 2009).
- 20 GAO, *Designing Evaluations*, p. 16.
- 21 European Commission, Directorate - General for the Budget, *Evaluating EU Activities: A Practical Guide for the Commission Services* (Luxembourg: Office for Official Publications of the European Communities, July 2004), ch. 3, pp. 37–43. European Commission, Financial Programming and Budgeting, Publications. http://ec.europa.eu/budget/index_en.htm (Apr. 27, 2009).
- 22 European Commission, *Evaluating EU Activities*, “Designing an Evaluation,” ch. 3, p. 45.

- 23 GAO, *Designing Evaluations*, p. 66.
- 24 These and the questions in the following paragraph are adapted from Valerie J. Caracelli and Judith A. Droitcour, "Evaluation and Audit Cross - discipline Models: A Multiplist Approach to Strengthening Information Quality," appendix I, table A1, p. 48, presented at the American Evaluation Association, Evaluation 2000, Conference History, Evaluation 2000, Session 277, Honolulu, Thursday November 2, 2000. www.eval.org (Apr. 27, 2009).
- 25 W. K. Kellogg Foundation, *W. K. Kellogg Foundation Evaluation Handbook* (Battle Creek, Mich.: Jan. 1, 1998), p. 59. [www.wkkf.org/Pubs/Tools/ Evaluation/Pub770.pdf](http://www.wkkf.org/Pubs/Tools/Evaluation/Pub770.pdf) (Apr. 27. 2009).
- 26 W. K. Kellogg Foundation, *W. K. Kellogg Foundation Evaluation Handbook*, pp. 59–60. Emphasis in original.

Chapter 4: Conducting Program Evaluation

An evaluation team's executing a program evaluation successfully means that it has incorporated in the design the various social and cultural perspectives of different groups' needs, values, and beliefs. The design is flexible enough that the team can redirect it as the project's conditions change. And the design and evaluation questions have been linked so that evaluators stay focused on objectives rather than methods. Program evaluations that have been carefully designed are also likely to reveal any inherent biases in the work, gather data only on relevant information, and lead to valid and reliable findings and interpretations.

This chapter points evaluators to some of the methodologies available to them. It also suggests that evaluators' gaining experience in conducting and reporting evaluations is critical in developing the evaluative expertise of the organization, as well as that of the larger evaluation community.

[4.1] Working through Evaluation Methodology

A program evaluation's methodology may be retrospective or prospective or both. Conducting program evaluations with a glance either backward to what has happened or forward to what can happen need not be complex, although it can be. Table 1 displays some aspects of these two methods.

As the evaluation team chooses its sources of data and methods of analysis, it should be sure to link them to expectations for the credibility and reliability of the findings. This will give greater assurance of the evaluation's quality.

As for specific methods and instruments of data collection and analysis available to evaluating organizations, table 1 indicates a handful of retrospective methods—

Table 1. Program Evaluation Methodology: Retrospective vs. Prospective

Method's feature	Retrospective method	Prospective method
Sources of questions	Criteria and issues relevant to policies, programs, or regulations	Ideas and assumptions about problems, probable causes, and possible solutions
Sources of information	Interviews, observations, opinion surveys, documents, administrative data	Prior research, theory, and evaluations; pilot or experimental tests of proposed approach; expert opinion
Types of analysis	Qualitative or quantitative approaches to empirical data; information synthesis on program criteria and issues	Simulations, models, and information syntheses related to conceptual and operational assumptions of proposals; Delphi techniques; analysis of likely effects

Source: Adapted from GAO, Prospective Evaluation Methods: The Prospective Evaluation Synthesis, PEMD - 10.1.10 (Washington, D.C. Nov. 1990), table 1.2, p. 16. www.gao.gov.

conducting interviews and surveys, for example, or examining documents or administrative data or conducting an information synthesis—and a handful of prospective methods—running pilots or experimental tests of a proposed approach, perhaps, or setting up simulations or models related to conceptual and operational assumptions. Many others are also available.

While performance audits conventionally provide information on how well things are going in a program, focusing on what can be readily measured and reported, program evaluations explore a program's performance more and why it is performing as it is and, perhaps, its possible future. Evaluations can supplement performance audit by examining relationships between a program's activities,

context, and results to help understand the reasons for the results and to differentiate the program's contribution to them from other factors.

Evaluators base their questions, and the types of evaluations they conduct, on the stage of the program they plan to study. Table 2 lists questions evaluators commonly ask, depending on a program's evolution, and it indicates the type of evaluation they are then likely to perform.

The purpose of this chapter, however, is not to list and define all possible methodologies; it is to offer SAIs resource opportunities for many different situations. The evaluation community provides a wealth of what are variously called checklists or cookbooks or matrixes or the like for selecting appropriate data sources and analysis methods, depending on the type of evaluation question.

A few methodological resources follow in the sections below. The first section contains direct resources from guides issued by the evaluation community and SAIs within that community. The second section includes indirect resources, those that derive from methodological discussions within evaluation reports. SAIs should aim for experience with a variety of methods, since no two evaluations are identical and since evaluation design may evolve as evaluators gain information about an evaluation's issues and data.

Table 2. Types of Questions and Evaluations by Program Stage

Program stage	Common evaluation questions	Evaluation type
Design	Is the program design well formulated, feasible, and likely to achieve the intended goals?	Design assessment
Early stage or new initiative within the program	<ul style="list-style-type: none"> • Is the program being delivered as intended to the targeted recipients? • Is the program well managed? • What progress has been made in implementing new provisions? 	Process or implementation evaluation
Mature and stable with a well-defined program model	<ul style="list-style-type: none"> • Are desired program outcomes obtained? • Did the program produce unintended side effects? • Do outcomes differ across program components, providers, or client subgroups? 	Outcome evaluation
	Why is the program no longer obtaining the desired outcomes?	Process evaluation
	Did the program cause the desired impact?	Impact evaluation

Source: Eric Bernholz and others, "Evaluation Dialogue between OMB Staff and Federal Evaluators: Digging a Bit Deeper into Evaluation Science," PowerPoint presentation [Washington, D.C.], July 2006, table 1, p. 14, and GAO, *Performance Measurement and Evaluation: Definitions and Relationships*, GAO - 05 - 739SP (Washington, D.C.: May 2005), www.gao.gov.

[4.1.a] Some Directly Available Methodology Resources

This paper has made reference to the *W. K. Kellogg Evaluation Handbook*. Part 2 of

the foundation's handbook, "Blueprint for Conducting Project - Level Evaluations," walks evaluative organizations through implementation step 5, "determining data - collection methods"; step 6, "collecting data"; and step 7, "analyzing and interpreting data."¹ The handbook contrasts quantitative and qualitative methods of analysis, describes techniques for conducting observations and interviews, illustrates the utility of tests and assessments, and suggests various types of document reviews. As it does so, it instructs evaluators in how to use these methods, gives examples of how others have used them, and lists aides to memory for implementing them. Step 7 also describes nontraditional forms of analysis, pointing up the goal of flexibility in evaluation design.

The online, continually updated, Evaluation Checklists Project of Western Michigan University's Evaluation Center currently offers some 32 files that give evaluators models of program evaluations and metaevaluations, discuss evaluation values and criteria, help evaluators manage evaluations, and aid them in evaluation capacity building. For example, the several checklists under "Metaevaluation" help in evaluating the performance of metaevaluations by scoring an evaluation's analysis of quantitative and qualitative information.²

USAID's EvalWeb, in the process of "re-energizing its central evaluation function," is drafting "Functional statements and position descriptions for the new evaluation division" and planning "a visible public 'launch' of the new evaluation unit early in 2009, along with creating "An initial agenda of studies for the central evaluation function."³ EvalWeb has long made available several evaluation resources, among them documents describing various evaluation methodologies such as participatory evaluations, key informant and focus group interviews, direct observation techniques, rapid appraisals, and impact assessments.

Gathered on USAID's Development Experience Clearinghouse pages, they include, for example, the 2006 *Moving from Outputs to Incomes: Practical Advice from Countries around the World*. Based on 2 days of discussion among members of courts of audit, finance ministries, treasury departments, and other national public service agencies from 12 nations in Africa, Europe and the United Kingdom, Latin America, the Middle East, and the United States, this evaluation resource identifies

what needs to be done in terms of both small steps and large steps that have had success in reorienting government systems to an outcome approach. This includes good practices, learning, and practical steps that could be of use to other governments in further developing an outcome - oriented approach to public sector management.⁴

For instance, under “Evaluation” in the chapter entitled “The Role of Monitoring and Evaluation in Making an Outcome Approach Possible,” *Moving from Outputs to Incomes: Practical Advice from Countries around the World* describes several purposes countries have found for evaluation. Among them are

- Assessing the appropriateness of programs, policies, and strategies for the present and the future.
- Identifying the continued appropriateness of objectives, indicators, and targets used for monitoring.
- Providing “why” and “how” information that is needed for an understanding of how and in what circumstances a program approach “works” or not.
- Explaining the factors behind positive impacts.
- Identifying unintended or unexpected consequences.
- Demonstrating causality or attribution— the linkage between inputs and activities and results.⁵

[12] RISK ANALYSIS—EVALUATION'S COSTS AND BENEFITS:

[S]ince evaluative resources in New Zealand are very limited, departments will need to make careful decisions about when to undertake evaluations Therefore, our analysis has considered how well officials decide what to evaluate, when to commission evaluative activity and what type of evaluative activity to undertake.

A central theme of the Review of the Centre report was the need to focus the public management system increasingly upon achieving outcomes for citizens. Consequently, our interest in evaluative activity is about how it can help improve decision-making, so that government interventions achieve better outcomes for citizens. We have therefore also looked at how effectively departments use the results of evaluative activity to inform policy and programme design and delivery.

Shenagh Gleisner, Arati Mascarenhas, and Sonia Wansbrough, "Improving Evaluative Activity in the New Zealand State Sector," paper presented at the Australasian Evaluation Society Conference, September 16–18, 2003, Auckland, p. 3. Australasian Evaluation Society, 2003 Conference, Paper WE31, "Government Cannot Make This Decision without a Government Study—Yeah, right." www.aes.asn.au (Apr. 29, 2009).

Much earlier, USAID's 1987 *AID Evaluation Handbook: AID Program Design and Evaluation Methodology Report 7* had set forth the agency's evaluation practices. Among the "sources of information and data for AID evaluations" in chapter 3's "Planning an Evaluation: What to Evaluate," are methods listed for collecting "secondary and contextual data" such as

- Host country development plans and policies
- Host country project records, reports

- Private sector organizations' reports
- Books, periodicals, and journals
- Research studies
- Other bilateral/multilateral donor project and program documentation
- Informant reports/expert opinions
- World Bank reports/studies

and for collecting primary data such as

Observation: Participant or non - participant . . . developed as part of regular site visits by project staff

Survey: Through interviews or by using structured questionnaires

Other:

- Case studies of before/after conditions
- Record-keeping by project staff in the form of journals, etc.
- Group sessions to stimulate discussion on project experience and lessons learned.⁶

Already noted is the European Commission's 2004 *Evaluating EU Activities: A Practical Guide for the Commission Services*, in which annex E thoughtfully describes "a number of tools and techniques often found in evaluation methods."⁷ For example, the Commission suggests that when evaluating organizations frame a design, they might consider intervention logic analysis or SWOT analysis of the strengths and weaknesses of the entity being evaluated and its future opportunities and threats.

Among retrospective and prospective methods the guide lists for data collection are interviews and focus groups. Annex E also specifically details different methods of analyzing qualitative and quantitative data, concluding with a discussion of cost - benefit analysis and multicriteria analysis, which "is more usually used within

the framework of prospective analyses and can be employed to support judgements in complex situations.”⁸ Additionally, chapter 1, “Evaluation in the Commission, Its Definition and Scope,” differentiates between retrospective and prospective methods under the heading “Temporal Variants.”⁹ Finally, the guide’s annex F briefly presents some of the methodological and analytical implications of evaluating policy, as distinct from evaluating agencies or programs.¹⁰

These are only a few of the guides, handbooks, and prescriptive resources that are available around the world, both electronically and on paper. Additional thoughts on evaluation methodology and its relationship to performance audit can be considered through a considerable body of technical literature reviewing and analyzing SAIs’ overall audit structure and practice.¹¹

[4.1.b] Some Indirectly Available Methodology Resources

Evaluation methodologies are also available in many evaluation reports. The description of an evaluation’s methodology, with discussion of its objectives, design, and scope, may appear within the main report with the main findings and conclusions, or it may be placed in an appendix to the report’s main body. Here are several instances of methodologies in published evaluations or in SAIs’ summaries of their evaluations.

In October 2004, after an evaluation of the research and evaluation function in four criminal justice agencies, New Zealand’s Controller and Auditor – General summarized the evaluation’s principal findings in a small document called Key Success Factors for Effective Co – ordination and Collaboration between Public Sector Agencies.¹² This was entirely separate from the main report, Co – ordination and Collaboration in the Criminal Justice Sector, which discussed the rationale for the evaluation, including its objectives and methodology.¹³ The purpose of the study had grown from the government’s perceived need for collaboration between public agencies with common goals. The criminal justice sector—a complex network of discrete but procedurally connected agencies—was chosen for evaluation because it

was “relatively well defined and self - contained,” had “a large amount of information available,” “functional relationships between the sector agencies [were] generally well understood,” and “agencies within the criminal justice sector often share[d] the same clients.”¹⁴

In the report’s section called “How Did We Carry Out the Audit?” the methodology was described as a series of activities that were and were not done. For example, New Zealand’s evaluation team examined policies and processes for coordination between the criminal justice sector agencies in four main areas, interviewed key officials, and examined pertinent documents, including legislative histories and “internal corporate policies relevant to strategic planning, policy development, information systems and responsiveness to Maori.” The team did not, however, “examine how criminal justice agencies co – ordinate their efforts and consult on day – to – day operational activities in relation to the management of individual clients and offenders.”¹⁵

The findings reported in Co – ordination and Collaboration in the Criminal Justice Sector included “many examples of good practice across the sector, and a strong commitment to sharing information and collaboration” but, at the same time, “the impact of one agency’s plans or activities on other agencies . . . had not always been well understood, creating risks for the completion of policy projects essential to meet Government strategic goals.”¹⁶ Among the study’s results were that “new processes [were] put in place for project oversight and managing risk” and a recommendation that the criminal justice sector consider a model that social policy agencies had adopted, since “The Social Policy Evaluation and Research Committee was established with a mandate to oversee the Government’s investment in social policy research and evaluation.”¹⁷ It might be possible, thus, to describe New Zealand’s work here as using a prospective method in an outcome evaluation.

[13] THE ELEMENTS OF A GOOD PERFORMANCE STORY

What is the context?

- the overall setting of the program (description, objectives, resources)
- the results chain (program theory)
- the risks faced

What was expected to be accomplished at what cost?

- statement of the (clear and concrete) outputs and outcomes expected
- planned spending

What was accomplished in light of these expectations?

- the relevant outputs delivered at what cost
- the outcomes realized related to the expectations
- a discussion of the evidence available demonstrating the contribution made by the program to those outcomes

What was learned and what will be done next?

- a discussion of what will be done differently as a result of what was achieved

What was done to assure quality data?

- a description of what the organization does to ensure the quality of the data and information reported

The main story line of a performance story is how well has the program performed in relation to what was expected and what will now be done differently to better ensure future performance.

John Mayne, Office of the Auditor General of Canada, Reporting on Outcomes: Setting Performance Expectations and Telling Performance Stories (Ottawa: May 14, 2003). Office of the Auditor General of Canada, Reports and Publications, Audit Methodology, Discussion Papers. www.oag-bvg.gc.ca (May 2, 2009).

Macedonia issued in July 2008 what might be called a process evaluation. In a presentation at a November 2008 EUROSAI seminar in Prague, representatives of the Macedonia State Audit Office noted that the principal issues in the study, called “Providing Proper Social Care,” had been to determine

- whether the family protection has proper legal framework;
- whether the Centers for social care have a proper human and other required resources to implement the legal regulations;
- whether there is a proper link and cooperation among the institutions involved in the social care system to ensure their efficient operation and to reduce the negative occurrences.¹⁸

The presentation outlined the study’s planning, data collection and analysis, and reporting methods, in addition to detailing its findings and recommendations. From site visits to eight social care centers in Skopje and other towns in the Republic of Macedonia, and from analysis of other primary and secondary data, the review had assessed whether the country’s family services program had a sound legal basis; staff were adequately trained and resourced; and the related school, mental health, police, and judicial institutions were sufficiently coordinated to ensure the provision of proper care to the families affected by the program.¹⁹

The evaluation’s findings had led to several recommendations that would adopt norms and operational standards for appropriate staffing of the institutions of the social care system, as well as their working space and equipment; develop a strategic budget plan for the operation of the social care centers and their oversight agencies; require that the agencies develop and implement methodological guidelines; strengthen the laws and regulations protecting and supporting victims of domestic violence; and enhance “the current forms of non-institutional care . . . that show good results [and] introduce new forms, based on the experience of the developed European countries.”²⁰

Additional recommendations would lead to the adoption of interagency protocols for more efficient cooperation between schools, social care centers, the police, health institutions, and the courts—as well as social workers, psychologists, teachers, and others with relevant expertise—in detecting and preventing adverse social

phenomena. Particular recommendations were made for training and for establishing family emergency services, counseling services, and other tools of intervention and support.²¹ Given that the review's results were geared toward greater success of the program, partly by adopting others' relevant experience, the evaluation might be said to have been prospective.

In January 2003, the Public Modernisation Programme for the Irish Public Service, having undertaken an evaluation of the national civil service, issued a 138-page final report, *NGM Report on Communicating Change and Modernisation in the Civil Service: A Survey of Irish Civil Servants*. The evaluation concluded, among other things, that civil servants had a positive attitude toward the need for change and modernization, that many felt that the existing program had not been adequately explained and was difficult to understand, and that there was a strong demand among all civil servants for more information about existing program initiatives. The result was a set of recommendations for developing "a formal communications strategy for the change and modernisation process" and improving "message structure, content and targeting in the communications process."²²

Like many program evaluation reports, the main body consisted of findings, conclusions, and recommendations, while an appendix explained the evaluation's methodology, clarifying the text with a number of data tables. In the methodology appendix, SAIs and others in the evaluation community can find the details of the target population for the survey, the evaluation team's sampling frame and sampling procedures, the survey response rates, and the evaluation team's methods of analysis.²³ The work might be called an outcome evaluation using retrospective methods.

The National Audit Office of Denmark, reorganized several years ago to enhance Rigsrevisionen's ability to do "cross-disciplinary examinations and recruit skills that would strengthen the performance audit activities and meet the challenges related to the introduction of electronic data processing," continually updates its guidelines, making its publications easily accessible on its Web site.²⁴ SAIs can also find there a small but useful book called *Performance Audit in Practice* whose purpose is to provide

snapshot pictures of how performance auditing in municipalities and counties

is carried out, and how subsequent reporting takes place. It is the Committee's objective to contribute to continued development in public sector auditing, so that it generates high quality and value for the individual municipality and county, regardless of who the auditor is and who is being audited.²⁵

The book explains that audits summarized in *Performance Audit in Practice* consist of two types: ongoing and extended performance audits. The basis of an ongoing performance audit is the “analyses and assessments carried out in support of the financial audit”; as a rule, this “highlights critical areas where a need for further examination exists.” An extended performance audit, in contrast, is an examination “of a more in-depth and/or broader nature” that “Normally provides the opportunity of answering critical questions.”²⁶

The section called “How is extended performance audit carried out?” notes in particular that

Extended performance audit is carried out independently from, but often as an extension of, the ongoing audit for example, if the audit highlights areas that warrant closer attention. More specifically, such examinations often concern management of selected areas, including

- Assessments of resource management
- Structural planning of work
- Calculation of financial latitude for use in political prioritisation
- Procedural analysis
- Assessments of the quality of case administration
- Assessments of target achievements and appropriateness of designated objectives.²⁷

The 15 summaries that follow this description include a wide range of studies, from residential facilities for child care to benchmarking between hospital departments to municipal payroll management to financial management of local fire service to information technology security policy. Several studies would appear to have used methods that were prospective in nature, in what might have been classified as

process, outcome, or impact evaluations.

The objective of study number 7, for example, was to map and assess organizational conditions regarding the administration and management of “an impending project concerning the amalgamation of [a] municipality’s utility companies into a common supply source.” The study’s methods included “interviews with relevant parties as well as [reviewing] financial data, function descriptions and other written materials.” The results included proposals to establish common supply, a draft for a chart of accounts structure with regard to common supply, and the auditors’ helping “to implement a strategic action plan for common supply.”²⁸

Similarly, the objective of study number 16 was to survey

the integration sector . . . to assess the municipality’s possibilities for controlling the correlation between Danish lessons and job creation schemes. For example, management possibilities were to be assessed according to both the current structure and alternative organisational forms.²⁹

Key personnel were interviewed, including the head of administration, integration employees, and management staff of the municipal language schools. In determining “Whether the current structure of the introduction process provided sufficient opportunity to realise the objective of correlation,” the evaluators finally presented to the municipality “A number of alternative structural models, as well as the financial advantages and disadvantages of the different models.”³⁰

Finally, models of evaluation design and methodology often surface in papers presented within the professional evaluation community at conferences, seminars, and workshops. To cite just one example, a Swedish paper, presented in 2000 at a session on citizen perspectives on evaluation at the European Evaluation Society conference in Lausanne, Switzerland, summarized the results of the study’s objectives: to “provide new information and knowledge to guide the development of the [Swedish] Tax Administration’s strategy and its business processes,” to “enhance its analytical skills and improve the taxpayer surveys as a means of data collection,” and to “test the usefulness of SEM [structural equations modeling] in this context.”³¹

The paper's methodology section then described the construction and analysis of statistical models that enabled the evaluator to answer several questions designed to meet the study's objectives:

The methodology offers a scientific tool to take measurement errors into account, to test models for goodness of fit, to give confidence intervals for explanatory variable's importance, to test group differences, to get feedback about possible improvements of proposed models etc. With a proper design, representative samples without too much nonresponse, the SEM-modelling creates visible relationships of true importance for linking citizens attitudes and behaviour with societal efforts to match their needs of service and at the same time live up to other functions. . . . In short, the statistical survey approach guarantees that every citizen has a fair chance to influence the modelling results, which is our way of taking evaluation to the people.³²

Clearly, much can be learned about evaluation methodology and its practice from both guidance and example.

[4.2] Framing Project Management Plans

Evaluating organizations will want to set up a framework for managing four phases, at minimum, in program evaluations: design, data collection, data analysis, and presentation of results. Each phase of an evaluation's management plan will have a set of constraints, bound by cost, time, location, and available staff and resources. Many models of such plans are available in many places. For example, in *Practical Evaluation for Public Managers*, chapter 9, called "What's Next?" walks readers easily through a number of suggested milestones, offered from experience in the Office of Evaluation and Inspections at the U.S. Department of Health and Human Services. As this department has done, SAIs can design their own evaluation process by phrasing such milestones as questions:

- How will potential study issues be identified? How will decisions be made about the issues to be examined? Who will be involved and

who will make the final decision?

- How will priorities be set?
- How will external stakeholders be involved, either in the planning or the performance of evaluations?
- At what stages will the [evaluation] team generate a product? Who will review it?
- Are formal comments solicited at any point? From whom?
- What kind of product results from the evaluation?
- Will the [evaluation] team make recommendations? To whom?
- Is there a mechanism to ensure the information is used?³³

Another, quite different model from Canada suggests that managing the reporting of outputs can evolve into reporting effects. From the perspective of performance audits, it offers, under the heading “Telling a Performance Story,” a challenging management metaphor:

Reporting on performance requires reporting on what was achieved in relation to what was expected, as well as on what was learned. . . . It involves more than simply reporting against several specific performance measures or targets previously set out. . . . It involves telling a credible performance story Indeed, it is good practice to ensure that the strategy for measuring program results includes room for addressing unintended results associated with the program.³⁴

While that discussion paper, *Reporting on Outcomes: Setting Performance Expectations and Telling Performance Stories*, is attuned primarily to the work of performance auditing, it suggests possibilities for alternative approaches to planning and reporting any evaluation. It also provides SAIs with a number of aids they can adapt in the service of managing program evaluation.

More traditional frameworks for working an evaluation management plan through to the published report abound, both on the Internet and in print. Since program evaluation is subsumed under performance audit in *Government Auditing Standards*, the “yellow book” that the U.S. Government Accountability Office issues, evaluating organizations might start with relevant passages on standards for field work and reporting in chapters 7 and 8, respectively. For example, when section 7.06 (in chapter 7) lays out that “Auditors must plan and document the planning of the work,” what is meant is that “Planning is a continuous process throughout the audit.”³⁵ The details of the audit include the project’s objectives, scope, methodology, and risk assessment, as well as matters of supervision, tests of evidence, and treatment of findings and documentation.

On the matter of reporting, chapter 8 advises, in section 8.04, that “Auditors should use a form of the audit report that is appropriate for its intended use and is in writing or in some other retrievable form. . . . The users’ needs will influence the form of the audit report.”³⁶ And, perhaps most importantly, reporting should include

the objectives, scope, and methodology . . . ; the audit results, including findings, conclusions, and recommendations, as appropriate; a statement about the auditors’ compliance with generally accepted government auditing standards; a summary of the views of responsible officials; and, if applicable, the nature of any confidential or sensitive information omitted.³⁷

SAIs may find a usefully concrete presentation on managing program evaluations in relevant sections of the European Commission’s *Evaluating EU Activities: A Practical Guide for the Commission Services*. This document, complete with charts and graphs, covers much of the ground regarding a project’s cost and time constraints, its development of qualitative and quantitative analysis plans, and the need to guarantee transparency and other quality factors in reporting evaluation results. For example, section 3.3.6 ties the evaluation work plan to the evaluation budget:

The work plan is always linked to the decision-making needs (i.e. the date by which the information from the evaluation is required . . .), and thus should never be drafted without this constraint in mind. When the link to the decision-making process has been established it is essential to establish clear

deadlines for different phases of the evaluation process, the meetings with the steering group and the deliverables to be supplied by the evaluator as the work progresses.³⁸

While the administrative constraints that the Commission notes appear to be unique to the European Union, the guide outlines a set of steps that could be useful, in principle, for managing any number of diverse program evaluations. For example, the guide lists the following among its several “evaluation process” steps:

Each evaluation shall be devised as a project comprising three separate stages: design, implementation and exploitation. The project manager shall establish who is to take part in these three stages . . .

When designing any (internal or external) evaluation project, the purpose of

[14] WRITING RESULTS:

Evaluation reports usually are written either to show progress toward reaching the stated goals and objectives and find areas that can be improved (a formative report) or to summarize the overall effects of the program (a summative report). The overall purpose of all reports is to communicate the effects of the program to the program staff, —clients of the . . . programs . . . , funding agency personnel, and the community at large. Unfortunately, many reports are sent to the funding agency for accountability purposes and are ignored by the other potential audiences. To some extent this may be the fault of professional evaluators who write in technical jargon without acknowledging the needs of various lay audiences or the program staff itself.

Judith Wilde and Suzanne Sockeye, *Evaluation Handbook* (Albuquerque: New Mexico Highlands University, Evaluation Assistance Center—Western Region, Dec. 1995), pp. 80–81. ERIC: Education Resources Information Center. Search Results www.eric.ed.gov (May 2, 2009).

the evaluation must be clearly and accurately defined. The evaluation project shall comprise the following points:

- a. the background to, reasons for and aims of the evaluation,
- b. for whom it is intended and who will use it,
- c. the scope of the area evaluated, d. the key questions,
- d. details of the information available,
- e. the reports,
- f. the deadlines.

In the event of external evaluation, the evaluation project must also include the contractual, financial and administrative clauses and quality criteria...³⁹

Above all, the organization managing successful program evaluations will ensure that evaluation staff have the greatest possible competence, knowledge, and skill for designing and executing evaluations and for sharing their experience with others in the organization and the larger evaluation community. Yet another evaluation handbook threads this thought throughout four chapters entitled “Thinking about Evaluation,” “Planning the Evaluation,” “Implementing the Evaluation,” and “Writing the Evaluation,” in themselves a rubric for the evaluative process.⁴⁰

Last is the drafting, designing, and production of the evaluation report. Managing this aspect of the project is no mean feat. Some helps to success that the evaluation team might follow include getting the people who want the project done and the people who will accomplish it to sit down and talk about it at the beginning, while ideas are fresh. The team might list all the necessary tasks in chronological order, making a master schedule with a deadline for each task and adjusting the schedule as the project changes.

[15] POLICY PERSPECTIVE ON EVALUATION CAPACITY:

Evaluative skills within both the State sector and the private sector to conduct high quality evaluative activity are relatively limited compared to existing demand.

. . .

. . . In order to make good decisions about what evaluative activity to undertake, advisors and decisionmakers need to understand what different types of evaluative activity offer at what cost and within what timeframe. However, people without this understanding often plan evaluative activity, without input from evaluators, resulting in poor decisions about what to evaluate. . . . In addition, many respondents expressed concern that policy staff generally have relatively poor evaluative and statistical analysis skills. . . . In particular, many policy staff reportedly expect too much certainty from evaluative findings

Shenagh Gleisner, Arati Mascarenhas, and Sonia Wansbrough, —Improving Evaluative Activity in the New Zealand State Sector, paper presented at the Australasian Evaluation Society Conference, September 16–18, 2003, Auckland, pp. 6–7. Australasian Evaluation Society, 2003 Conference, Paper WE31, —Government Cannot Make This Decision without a Government Study—Yeah, right. www.aes.asn.au (May 2, 2009).

The team members who write the document might plan the document, carefully organizing its files, whether paper or electronic, and maintaining them in a manageable size. Source materials can be separated by section (frontmatter, chapters, appendixes) and grouped by elements (text, tables, visuals), all with backup copies. In drafting the report, evaluators might write from the readers' perspective and actively, not passively, and in terms laypersons can follow, not in jargon. The team might also coordinate its work with the work of the final product's

designers and printers, not only for visual effect but also for timeliness. Finally, for quality control, everyone who can might proofread everything in the report that its readers will read. Transparency in evaluation design, evidence, findings, and conclusions can well extend to transparency of expression and presentation.

- 1 W. K. Kellogg Foundation, *W. K. Kellogg Foundation Evaluation Handbook* (Battle Creek, Mich.: Jan. 1, 1998), pp. 69–95. www.wkcf.org/Pubs/Tools/Evaluation/Pub770.pdf (Apr. 29, 2009).
- 2 Evaluation Checklists, The Evaluation Center, Western Michigan University, Kalamazoo, Michigan. www.wmich.edu/evalctr/checklists (Apr. 29, 2009).
- 3 USAID, EvalWeb, <http://evalweb.usaid.gov> (Apr. 29, 2009).
- 4 Burt Perrin, *Moving from Outputs to Outcomes: Practical Advice from Countries around the World* (Washington, D.C.: IBM Center for The Business of Government, January 2006), p. 5. USAID, EvalWeb, Evaluation Resources, <http://evalweb.usaid.gov> (Apr. 29, 2009).
- 5 Burt Perrin, *Moving from Outputs to Outcomes*, p. 43. (Apr. 29, 2009).
- 6 See USAID’s EvalWeb *AID Evaluation Handbook: AID Program Design and Evaluation Methodology Report 7* (Washington, D.C.: April 1987), p. 25. See also p.6, which “defines evaluation as a management activity that is undertaken selectively to inform managers about key issues before major decisions are made regarding . . . the issues of relevance, effectiveness, efficiency, impact, and sustainability.” USAID Development Partners Resources, USAID Evaluation Publications, USAID Development Experience Clearinghouse, Advanced Search. <http://dec.usaid.gov/partners/evalweb/index.cfm#> (Apr. 29, 2009).
- 7 European Commission, Directorate-General for the Budget. *Evaluating EU Activities: A Practical Guide for the Commission Services* (Luxembourg: Office for Official Publications of the European Communities, July 2004), annex E, p. 85. European Commission, Financial Programming and Budgeting, Publications. http://ec.europa.eu/budget/index_en.htm (Apr. 29, 2009).

- 8 European Commission, *Evaluating EU Activities*, pp. 85–90.
- 9 European Commission, *Evaluating EU Activities*, pp. 12–13.
- 10 European Commission, *Evaluating EU Activities*, annex F, pp. 91–96.
- 11 See, for example, Yutaka Suzuki, “Basic Structure of Government Auditing by a Supreme Audit Institution,” *Government Auditing Review 11* (March 2004): 39–53, esp. figs. 1 and 2. www.jbaudit.go.jp/pr/pdf/e11d03.pdf, as well as the extensive 2007 bibliography at www.iiasiisa.be/egpa/aggroup/Bibliography%20on%20performance%20auditing-july2007.pdf (Apr. 29, 2009).
- 12 K. B. Brady, Controller and Auditor-General, *Key Success Factors for Effective Co-ordination and Collaboration between Public Sector Agencies*, Report of the Controller and Auditor-General, Tumuaki o te Mana Arotake (Wellington: New Zealand Audit Office, Oct. 2004). New Zealand, Controller and Auditor-General, Good Practice Guides. www.oag.govt.nz (Apr. 30, 2009).
- 13 K. B. Brady, Controller and Auditor-General, *Co-ordination and Collaboration in the Criminal Justice Sector*, Report of the Controller and Auditor-General, Tumuaki o te Mana Arotake (Wellington: Oct. 1, 2003). New Zealand, Controller and Auditor-General, Reports by Type, Performance Audit Reports. www.oag.govt.nz (Apr. 30, 2009).
- 14 K. B. Brady, Controller and Auditor-General, *Co-ordination and Collaboration in the Criminal Justice Sector*, pp. 10–11.
- 15 K. B. Brady, Controller and Auditor-General, *Co-ordination and Collaboration in the Criminal Justice Sector*, p. 14.
- 16 K. B. Brady, Controller and Auditor-General, *Co-ordination and Collaboration in the Criminal Justice Sector*, p. 19.
- 17 K. B. Brady, Controller and Auditor-General, *Co-ordination and Collaboration in the Criminal Justice Sector*, pp. 19 and 51.
- 18 Biljana Ivanovska, State Audit Office, Republic of Macedonia, “Providing Proper Social

- Care,” presentation at EUROSAI Training Event, Seminar on Auditing of Social Security Systems, Prague, November 10, 2008, p.3. See Czech Republic, Supreme Audit Office, International Cooperation, Conferences [and] Seminars, 2008, EUROSAI Training Event, Seminar on Auditing of Social Security Systems, Prague, November 10–12, 2008. www.nku.cz/default-en.htm (Apr. 30, 2009).
- 19 Biljana Ivanovska, State Audit Office, Republic of Macedonia, “Providing Proper Social Care,” p. 5.
- 20 Biljana Ivanovska, State Audit Office, Republic of Macedonia, “Providing Proper Social Care,” pp. 14– 15.
- 21 Biljana Ivanovska, State Audit Office, Republic of Macedonia, “Providing Proper Social Care,” pp. 16– 17.
- 22 NGM Market Research and Strategic Planning, *NGM Report on Communicating Change and Modernisation in the Civil Service: A Survey of Irish Civil Servants*, report prepared for Civil Service Communications Strategy Group (Waterford, Ireland: Jan. 22, 2003), pp. 5–6. Publications, View Complete List of Publications Organized by Thematic Areas, Quality Customer Service. www.bettergov.ie/eng/index.asp (May 2, 2009).
- 23 NGM Market Research and Strategic Planning, *NGM Report on Communicating Change and Modernisation in the Civil Service*, sec. 3.0, app. 1, Methodology, pp. 70–75.
- 24 Denmark, Audit of the State Accounts, About Us, History and Background. www.rigsrevisionen.dk (May 2, 2009).
- 25 Denmark, Audit of the State Accounts, Committee for Public Audit, *Performance Audit in Practice: Examples from County and Municipality Audits* (Copenhagen: Jan. 2003), p. 5. Denmark, Rigsrevisionen, Performance Audit, Methodology and Planning. www.rigsrevisionen.dk (May 2, 2009).
- 26 Denmark, Audit of the State Accounts, Committee for Public Audit, *Performance Audit in Practice*, table 1, p. 6.
- 27 Denmark, Audit of the State Accounts, Committee for Public Audit, *Performance Audit in Practice*, p. 11.

- 28 Denmark, Audit of the State Accounts, Committee for Public Audit, *Performance Audit in Practice*, pp . 11–12.
- 29 Denmark, Audit of the State Accounts, Committee for Public Audit, *Performance Audit in Practice*, p.18.
- 30 Denmark, Audit of the State Accounts, Committee for Public Audit, *Performance Audit in Practice*, p. 18.
- 31 Maria Karanta and others, “A Citizen’s Perspective on Public Sector Performance and Service Delivery: Progress in Measurement and Modelling of Data from a Swedish Taxpayer Survey,” p. 3, paper presented at the European Evaluation Society conference, Lausanne, Switzerland, October 12–14, 2000. Canadian Evaluation Society, Grey literature. <http://evaluationcanada.ca/site.cgi?s=1> (May 2, 2009).
- 32 Maria Karanta and others, “A Citizen’s Perspective on Public Sector Performance and Service Delivery,” p. 6.
- 33 June Gibbs Brown, *Practical Evaluation for Public Managers: Getting the Information You Need* (Washington, D.C.: U.S. Department of Health and Human Services, Office of Inspector General, Nov. 1994), pp. 76–77. www.epa.gov/evaluate/DHHSPracticalEvaluation.pdf (May 2, 2009).
- 34 John Mayne, Office of the Auditor General of Canada, *Reporting on Outcomes: Setting Performance Expectations and Telling Performance Stories* (Ottawa: May 14, 2003). Office of the Auditor General of Canada, Reports and Publications, Audit Methodology, Discussion Papers. www.oag-bvg.gc.ca (May 2, 2009).
- 35 Comptroller General of the United States, *Government Auditing Standards, July 2007 Revision*, GAO-07-731G (Washington, D.C.: U.S. Government Accountability Office, July 2007), p. 124. www.gao.gov (May 2, 2009).
- 36 Comptroller General of the United States, *Government Auditing Standards, 2007 Revision*, p. 160.
- 37 Comptroller General of the United States, *Government Auditing Standards, 2007 Revision*, p. 161.

- 38 European Commission, *Evaluating EU Activities*, p. 46.
- 39 European Commission, *Evaluating EU Activities*, p. 69.
- 40 Judith Wilde and Suzanne Sockey, *Evaluation Handbook* (Albuquerque: New Mexico Highlands University, Evaluation Assistance Center–Western Region, Dec. 1995). ERIC: Education Resources Information Center, Search Results. www.eric.ed.gov(May 2, 2009).

*C*hapter 5: *D*eveloping *O*rganizational *C*apacity for *P*rogram *E*valuation

Generally speaking, no one evaluator has the range of skills necessary to conduct a full evaluation, and often evaluative organizations do not have the full expertise necessary to execute the seemingly vast array of possible evaluation designs. Evaluators may work in teams to combine their abilities and evaluative organizations both recruit evaluators into new staff positions and contract with external evaluators for the use of temporary expertise. It might be thought that working in teams and with external experts would help in conducting evaluations and would also bring new knowledge and skills to individuals and the organization in preparation for further evaluative work.

[5.1] The Evaluation Team

Whether and how well individual evaluators and their organizations learn from evaluations has been discussed at length. For example, at the European Evaluation Society's 2002 conference, at a session on facilitating the use of evaluations and what is learned from them, the opinion was expressed that "it is usually organized entities and not individuals that are expected to learn from the evaluations, which, of course raises special issues about how organizations use information and learn."¹

[16] DEVELOPING AN EVALUATION TRADITION:

In 1998 there were 9 national and regional evaluation associations, and by the fall of 1999 there were 26. . . . In recent months, several new associations in western Europe have incorporated; Central Eastern Europe is rapidly forming its own national evaluation societies; Russia has organized an evaluation network; the very successful 1999 African evaluation conference has spawned not only the African Evaluation Association, but also numerous national associations; several associations have been chartered in South and Southeast Asia; the new Japanese Evaluation Association has been formed; and plans are under way for new evaluation associations in the Caribbean and Latin America.

. . . the development of national and regional evaluation associations is critical to the field and profession of evaluation because evaluation associations bring together evaluators from multiple disciplines to share knowledge and experiences, bridge disciplinary divides, debate issues of fundamental importance, set standards and ethical guidelines, build skills, and chart the future as a group with a strong and shared identity. . . . Likewise, if we view institutionalizing evaluation as desirable, the probability that skilled evaluators are recruited into staff positions also reflects evaluator supply and a strong evaluation tradition.

A. J. Love and C. Russon, —Building a Worldwide Evaluation Community: Past, Present, and Future, *Evaluation and Program Planning* 23 (2000): 449–59. (May 2, 2009).

At the same conference, at a session on evaluation and organizational change, a more theoretical opinion was expressed:

Evaluation has evolved in a way that is conducive to transaction and aims at learning. Evaluation plays the role of exposing information and knowledge to the organization in order to make that knowledge actionable. Thus evaluation

serves an important function in building learning organizations by enabling them to understand, create and preserve their own organizational knowledge. . . . The evaluator . . . draws on her knowledge of participatory types of evaluation, types of knowledge, types of learning, components of organizational learning, to enable the organization . . . to become a learning organization in its own right.²

Moreover, precisely who constitutes an evaluation team has been discussed to some degree. In designing program evaluations, evaluators collaborate with requesters and other stakeholders to transform objectives into questions they can answer within the evaluation plan's constraints. To this extent, participation from others aids the evaluation team in framing a design.

Another paper at an earlier European Evaluation Conference in 2000, in a session on evaluation as a tool for civil society, touched on this topic, stating that "the current challenge for evaluation in civil society is building sufficient capacity to enable a wide range of stakeholders to conduct and use evaluation" and that "involvement in conducting evaluation not only builds organizational capacity for development and accountability, but also builds social capital in communities."³

The American Evaluation Association's *AEA's Guiding Principles for Evaluators* indicate that evaluators should ensure that an evaluation team has the education, skills, and experience necessary to determine what an evaluation requires; insight to know whether they can do the evaluation; and the capacity to work with other evaluators to learn from them. However, the principles also indicate, first, that

Evaluation is a profession composed of persons with varying interests, potentially encompassing but not limited to the evaluation of programs, products, personnel, policy, performance, proposals, technology, research, theory, and even of evaluation itself.

and, again, that given evaluators'

differences in training, experience, and work settings, the profession of evaluation encompasses diverse perceptions about the primary purpose of evaluation. These include but are not limited to . . . bettering products,

personnel, programs, organizations, governments, consumers and the public interest; contributing to informed decision making and more enlightened change; precipitating needed change; empowering all stakeholders by

[17] EVALUATION: A French Evaluation Society

Société Française de l'Évaluation (SFE)

Why? "To contribute to the development of the evaluation and to promote its use within public and private organizations".

For whom? "The SFE wants to be a meeting, debate, and training place open to all the evaluation stakeholders and users, regardless of whether they are elected officials, government employees, civil servants, researchers, academics, consultants, internal auditors or any other public or private sector evaluation professionals".

How? "It sets for itself the objective to foster the techniques and methods and to favor compliance with the rules of ethics and proper procedure in order to guarantee the quality of its evaluations along with the proper use of their results".

French Evaluation Society (SFE), Evaluation. www.sfe.asso.fr
(May 3, 2009).

collecting data from them and engaging them in the evaluation process; and experiencing the excitement of new insights.⁴

The National Science Foundation (NSF) in the United States has addressed where to find such professionals, if they are not already members of the evaluation organization's staff. NSF suggests contacting college and university department chairs, identifying independent contractors or consultants, or calling for recommendations from other researchers or evaluators, including those named in published evaluation reports.⁵ Assembling an evaluation team, NSF advises,

is, of course, a critical stage in the evaluation process. At the outset, the

sociocultural context in which the programs or projects are based must be taken into account. . . . At the very least, the evaluator or evaluation team should be fully aware of and responsive to the participants' and stakeholders' culture, particularly as it relates to and influences the program.⁶

Other organizations also describe skills and expertise evaluators should have. For example, the Swiss Evaluation Society has developed evaluation standards for promoting the quality of evaluation.

According to SEVAL, "Those who conduct evaluations are both competent and trustworthy [to] help ensure the results an evaluation reaches are accorded the highest degree of acceptance and credibility possible." Evaluators are able to "encourage stakeholders both to follow the evaluation process and to use the evaluation." They have the skills to identify, document, and communicate an evaluation's goals and procedures. They know "how to specify which information is indispensable to answer the questions posed by or in the evaluation, and [how] to distinguish it from information that is . . . desirable to know but ultimately unnecessary."⁷

Other standards evaluators meet, according to SEVAL, are that they are able to "utilize theoretical models and value orientations" to interpret "the information gathered, as well as the results." They can use statistical methods and can determine the contextual factors that need to be "closely investigated so as to appropriately situate the planning, execution, and communication of the evaluation." They know how to conduct evaluations "in a . . . cost-conscious manner," having the expertise to choose "the methods and procedures [that] are as practicable as possible." They are able to "assess and present the strengths and weaknesses that exist in the object being evaluated, in a manner that strengths can be built upon and problem areas addressed."⁸

[5.2] Developing Organizational Expertise

It has been suggested that "thinking of evaluation [is] everyone's responsibility" and that, in fact,

An important part of an evaluator's job (internal or external) is to assist in building the skills, knowledge, and abilities of other staff and stakeholders. It is better to have an evaluator who has spent time working with staff to integrate evaluation activities into day-to-day project management and delivery, than to have one who has conducted a perfectly constructed evaluation with strong recommendations that no one uses and with no one able to continue the work.⁹

Case studies of federal agencies in the United States have demonstrated something similar. For example, GAO's report *Program Evaluation: An Evaluation Culture and Collaborative Partnerships Help Build Agency Capacity*, concluded, in part, that

ACF's [Administration for Children and Families] long-standing collaborative relationship with ASPE [Assistant Secretary for Planning and Evaluation] helped build the agency's expertise directly—through advising on specific evaluations, as well as indirectly—through building the expertise of the research community that conducts those evaluations. ASPE coordinates and consults on evaluations conducted throughout HHS [the U.S. Department of Health and Human Services]. ACF staff described getting intellectual support from ASPE—as well as sharing in joint decisions and pooling dollar resources— which boosted the credibility of their work in ACF. At ACF, skills in statistics or research are not enough. They also require people with good communication skills, who can explain the benefits of participation in evaluations to states and localities. For decades, ASPE has funded evaluations, as well as research on poverty, by academic researchers, contract firms, and state agencies. ASPE staff described their investment in poverty research as providing additional assets for evaluation capacity because, in the field of poverty research, the academic world overlaps with the contract firms. They believe this means that (1) better research gets done because prominent economists and sociologists are involved and (2) research on poverty is better integrated with policy analysis than in other fields.¹⁰

An evaluative organization's evaluation teams should be able to demonstrate content knowledge, professional competence, and evaluative expertise. An organization that is conducting program evaluations with a staff whose members are successfully planning, designing, and managing evaluations, and advising one another on

[18] AUDIT AND EVALUATION AS COMPLEMENTARY:

This paper . . . examin[es] how audit and evaluation approaches as cross-discipline models provide useful information for programs and policy. GAO, and other professional services organizations focused on accountability, perform a variety of studies. Typically, these studies address questions in one or both of the following categories: (1) Audit-emphasis questions involving financial auditing, compliance, management controls, and agencies' attainment of pre-set goals and (2) evaluation-emphasis questions involving the development of new criteria, assessment of impact or net effect and other outcomes (including unintended outcomes), examination of technical soundness of government information (beyond completeness, accuracy, and consistency), and prospective analyses. . . . Audit and evaluation studies can bring different facets of information together, furnishing more comprehensive information on performance.

Valerie J. Caracelli and Judith Droitcour, U.S. General Accounting Office, —Evaluation and Audit Cross-discipline Models: A Multiplist Approach to Strengthening Information Quality, paper presented at Evaluation 2000, Session 277, American Evaluation Association Conference, Honolulu, November 2, 2000. American Evaluation Association, Training, Conference History, www.eval.org (May 5, 2009).

evaluation's concepts and practices, is an organization that is also developing and expanding its evaluative expertise.

Those who manage evaluations, and those who manage evaluators, will improve their understanding of key evaluation concepts, evaluation design, methodology, reporting, and evaluation assessment by interacting meaningfully with evaluators as evaluations are executed. Whether evaluations are developed and implemented internally within the evaluating organization or are conducted with assistance from

external contracts, the evaluation process entails a feedback loop: The evaluation results get translated into usable information for those they were intended for, and the evaluation process gets translated into an expansion of expertise for the evaluators and their organization.¹¹

[5.3] The Evaluation Community

WWW Virtual Library: Evaluation, The World Wide Evaluation Information Gateway is an Internet directory for evaluators linking to evaluators' communities, education and training sites, public and private evaluation organizations, and various resources for methods and analysis.¹² Under "Popular Sites" is a link to Free Resources for Program Evaluation and Social Research Methods, the top-rated site in this category and a prime example of what is available within today's diverse evaluation community. Created and maintained by evaluation experts, it focuses on methods and theory in evaluation and social research but also provides a wealth of information on "human research protection practices, research codes and guidelines, rights of the public, [and] professional standards and ethics"¹³

It links not only to the most sophisticated and academic of materials but also to beginners' guides for those who have little or no technical background. It has links to the world's courts of audit and to hundreds of evaluation societies and research centers in the United States, Europe, and the rest of the world. Besides linking to professional journals from A to Z, it shows how to access online discussion groups and mailing lists, where one may get information on how to resolve methodological problems, among others.

Besides the WWW Virtual Library, the International Organisation for Cooperation in Evaluation gives evaluators access to the increasingly large global evaluation community. Some 50 national and regional evaluation networks and societies can be reached directly through links to their Web sites from IOCE. News of worldwide evaluators' conferences and meetings are posted at IOCE, as are announcements of the formation of new evaluation societies in various countries. Case studies on how 14 different evaluation associations developed and operate their organizations in

Africa, the Americas, Asia, Australasia, and Europe are available at this site as well.¹⁴

But what of actual individuals in actual evaluation? SAIs might consider an evaluation like GAO's 2003 *Program Evaluation: An Evaluation Culture and Collaborative Partnerships Help Build Agency Capacity*. The evaluative organization's and evaluated agencies' evaluators came together, as described in a passage like this:

Four main elements of evaluation capacity were apparent across the diverse array of agencies we reviewed, although they took varied forms. These elements include an evaluation culture, data quality, analytic expertise, and collaborative partnerships. Agencies demonstrated an evaluation culture through commitment to self examination and learning through experimentation. Data quality and analytic expertise were key to ensuring the credibility of evaluation results and conclusions. Agency collaboration with federal and other program partners helped leverage resources and expertise for evaluation.¹⁵

GAO's report then described the evaluation culture that its team found among the case study evaluators:

Three of our cases—ACF [Administration for Children and Families], NHTSA [National Highway Traffic Safety Administration], and NSF—clearly evidenced an evaluation culture: they had a formal, regular process in place to plan, execute, and use information from evaluations. They described a commitment to learning through analysis and experimentation. HUD [the U.S. Department of Housing and Urban Development] and the Coast Guard had more ad hoc arrangements in place when questions about specific initiatives or issues created the demand for evaluations. HUD officials described an annual, consultative process to decide which studies to undertake within budgeted resources.¹⁶

Beyond the institutions' boundaries are the community of working evaluators, such as the more than 300 evaluators from 35 countries— some of them staff members in ministries of finance and similar public agencies—who participated in the

[19] COLLABORATION IN THE EVALUATION COMMUNITY:

Program evaluation started in the US with the Great Society. What was created then [were] new ways of acting to tackle social problems, and a new practice, that of evaluation . . . but . . . positivist and constructivist paradigms opposed each other . . . [W]e are still waging a rearguard war between qualitative and quantitative methods.

. . .

However, . . . principal and agent . . . should share goals . . . Both . . . should contribute to program elaboration, goal definition, etc. They should understand that they can both benefit from evaluation . . . if there were a continuous interaction between lower and higher levels of the hierarchy. . . .

. . . At the lower level, if people knew what they were doing, what to expect, and how they could contribute to the outcome, then they would be more favourably disposed to evaluation, and understand that it is for their good. At the higher level, theories received from below would be more grounded, and better suited to understanding a complex and diverse situation.

Nicoletta Stame, —Evaluation and the Policy Context: The European Experience, pp. 37-39 and 42, keynote address, Australasian Evaluation Society 2003 conference, Auckland, November 14- 18, 2003; see *Evaluation Journal of Australasia*, New Series, 3:2 (December 2003): 36-43. Australasian Evaluation Society, Publications. www.aes.asn.au/publications (May 4, 2009).

September 1999 African Evaluation Association conference in Nairobi. The subject, “Increasing Evaluation Capacity in Africa,” produced some 80 papers on all major areas of evaluation research.

Another working group is the European Evaluation Society. Like a number of other such professional evaluation associations, its stated purpose is to promote the theory, practice, and use of program evaluation by bringing professional practitioners together through seminars, biennial conferences, training and work opportunities, and other forms of networking. On EES’s Web site, for example, is the 2009 news that besides the free online availability of a new book on country-led monitoring and evaluation systems, issued by UNICEF in partnership with the World Bank, UN Economic Commission for Europe, IDEAS (International Development Evaluation Association), IOCE, and others, in just the first 6 months of the year evaluators’ conferences were to be, or had been, conducted in

- Berlin, on global health care;
- Brussels, on the evaluation of humanitarian action;
- Vukovar, Croatia, on empowering the community;
- Potsdam, on policy evaluation in the European Union; and
- Southern California, at Claremont Graduate University, on evaluation and applied research methods.¹⁷

SAIs who seek to foster a culture that includes program evaluation in their organizations will find themselves operating in a new and different environment—perhaps even, as was suggested in chapter 1, with new beliefs and values. The Cour des Comptes of Morocco has made this distinction, also, noting that the task of audit institutions is to clarify their role and responsibility as “guardians of the integrity, legality and efficiency of public spending,” whereas “there are constraints to be faced when joining up an evaluation with an audit,” since “evaluation often serves to improve existing programs” rather than leading to fundamental changes in the allocation of resources. Further, “A specific evaluation may even lead to suggesting

[20] ASSESSING EVALUATION—CONTROL VS. COOPERATION:

In the context of management reforms, Supreme Audit Institutions have been more concerned to assess and report their own performance However . . . there are several difficulties . . . when evaluating performance audit effectiveness. For example, it is not guaranteed that accepting the auditor's recommendations really leads to a better performance. In addition, in order for the auditee to accept the auditor's recommendations, the relationships with the auditees should become more important to the auditors, as a result of which the auditors become more attached to the daily work of the audited body. But this may threaten the auditors' and the SAI's independence and objectivity.

. . .

The connection between the auditor's role and how the audited body perceives it was not proven [in this study]. However, the present study reveals that the audited body recognized the auditor's consulting role, based on joint co-operation, more positively than any watchdog role implying a form of control.

Jane Etverk, — Measuring Performance Audit Effectiveness: The Case of Estonia, M.A. thesis, Department of Public Administration, University of Tartu, Tartu, Estonia, June 27, 2002, pp. 46- 48. National Audit Office of Estonia, About NAO, Documents, Other Documents, Tallinn, Estonia. www.riigikontroll.ee/?lang=en (Apr. 17, 2009).

that a program remain unchanged.”¹⁸

In the final analysis, the question is: How do evaluative organizations build a capacity for evaluation, and how do they enter the evaluation community? One of the more basic answers was offered by the Inspector General for the U.S. Department of Health and Human Services Office of Evaluation and Inspections.

She said,

[W]e often must assess and evaluate programs . . . [with] methods consistent with principles of evaluation but quite different from the exacting work of scientists . . . We see our approach as far more attainable and our goals as very much like those of most program managers. We examine and report on issues of efficiency, fraud and abuse, and program effectiveness. . . . Like many program managers, we move quickly from one issue to the next as we anticipate and respond to the interests and needs of decisionmakers in the executive and legislative branches of government.

We are interested in making our commonsense approach to evaluation – providing outcome – oriented, practical information for immediate use – available to program managers and others who may not consider themselves evaluators. We realize that establishing a distinct evaluation component like ours may be an unreachable or even inappropriate goal for many programs. It requires the help of experienced evaluators, a long – term commitment, and more than a little risk – taking. It’s hard work. However, with an understanding of the way evaluation techniques can apply to program management, non – evaluators can effectively use the services of evaluators to obtain needed information. Similar to any relationship based on a contract for services, the more you know about the concepts and applications of evaluation, the stronger your partnership with professional evaluators will be, and more likely you will be to end up with information that meets your needs.¹⁹

In brief, a local fellowship may also be an extended network of experts from around the world.

1 Barbrol Anell and Marja E. Lemne, “Learning from Evaluations: Who Learns What, When, How and Why?” paper presented at the European Evaluation Society 5th biennial conference, Seville, October 2002, p. 7. Canadian Evaluation Society, Grey literature. <http://evaluationcanada.ca/site.cgi>. (May 3, 2009).

- 2 Barbara Rosenstein and Miri Levin - Rozalis, "Program Evaluation and Organizational Learning: A Theoretical Perspective," paper presented at the European Evaluation Society 5th biennial conference, Seville, October 2002, p. 17. Canadian Evaluation Society, Grey literature. <http://evaluationcanada.ca/site.cgi> (May 3, 2009).
- 3 Mark Bitel, "Taking Evaluation to the People? Who Wants It?!" p. 1, paper presented at the European Evaluation Society 4th biennial conference, Lausanne, October 2000. Canadian Evaluation Society, Grey literature. <http://evaluationcanada.ca/site.cgi> (May 3, 2009).
- 4 American Evaluation Association, *Guiding Principles for Evaluators* (Fairhaven, Mass.: July 2004), p.1. American Evaluation Association, Publications. www.eval.org (May 3, 2009).
- 5 John Frechtling, *The 2002 User Friendly Handbook for Project Evaluation*, NSF 02 - 057 (Arlington, Va.: National Science Foundation, Jan. 2002), app. A, p. 84. National Science Foundation, Publications, Search/Browse Publications. www.nsf.gov/index.jsp (May 3, 2009).
- 6 John Frechtling, *The 2002 User Friendly Handbook for Project Evaluation*, ch. 7, p. 65.
- 7 Thomas Widmer, Charles Landert, and Nicole Bachmann, *Evaluation Standards of Seval, The Swiss Evaluation Society (SEVAL Standards)* (Fribourg: Swiss Evaluation Society, Dec. 5, 2000), pp. 4–6. Swiss Evaluation Society, Seval Standards. www.seval.ch/en (May 3, 2009).
- 8 SEVAL Standards, pp. 6–10.
- 9 W. K. Kellogg Foundation, *W. K. Kellogg Foundation Evaluation Handbook* (Battle Creek, Mich.: Jan. 1, 1998), p. 68. www.wkkf.org/Pubs/Tools/Evaluation/Pub770.pdf (May 3, 2009).
- 10 GAO, *Program Evaluation: An Evaluation Culture and Collaborative Partnerships Help Build Agency Capacity*, GAO - 03 - 454 (Washington, D.C.: May 2, 2003), p. 20. www.gao.gov (May 3, 2009).
- 11 See, for example, the related discussion in European Commission, *Evaluating EU Activities*, European Commission, Directorate - General for the Budget, *Evaluating EU*

Activities: A Practical Guide for the Commission Services (Luxembourg: Office for Official Publications of the European Communities, July 2004), pp. 25–27. European Commission, Financial Programming and Budgeting, Publications. http://ec.europa.eu/budget/index_en.htm (May 3, 2009).

- 12 WWW Virtual Library: Evaluation, The World Wide Evaluation Information Gateway, www.policy-evaluation.org (May 3, 2009).
- 13 Free Resources for Program Evaluation and Social Research Methods, <http://gsociology.icaap.org/methods> [May 4, 2009].
- 14 IOCE, International Organisation for Cooperation in Evaluation. <http://ioce.net/index.shtml> (May 4, 2009).
- 15 GAO, *Program Evaluation: An Evaluation Culture and Collaborative Partnerships Help Build Agency Capacity*, GAO - 03 - 454 (Washington, D.C.: May 2, 2003), p. 9. www.gao.gov (May 4, 2009).
- 16 GAO, *Program Evaluation: An Evaluation Culture and Collaborative Partnerships*, pp. 9–10.
- 17 European Evaluation Society, News. www.europeanevaluation.org (May 5, 2009).
- 18 Kingdom of Morocco, Cour des Comptes, “Suggestions and Comments on the Draft Report of the INTOSAI Working Group on Program Evaluation,” September 2005, p. 3.
- 19 June Gibbs Brown, *Practical Evaluation for Public Managers: Getting the Information You Need* (Washington, D.C.: U.S. Department of Health and Human Services, Office of Inspector General, Nov. 1994), pp. iii–iv. www.epa.gov/evaluate/DHHSPacticalEvaluation.pdf (May 3, 2009).

Chapter 6: Perspectives on the Future

What does program evaluation have in store for the relationships SAI's have with their own audit and evaluation societies, INTOSAI work groups, and public and private evaluative institutions? What is their role in developing evaluation capacity among individuals and within the evaluation community at large? What is in store for the future?

[6.1] SAI's as Agents of Change

Evaluation research has noted that with the growth of the quantity and size of performance audits, state auditors have tended to move toward managing accountability through the evaluation of performance. Such change has sometimes been accompanied by the expectation that SAI's will alter the way reform is implemented or that they will serve as agents of managerial change, even though this contrasts with their traditional role. Often, it has been noted,

SAI's are still expected to execute a control role: to act like watchdogs on behalf of the public But in acting as a watchdog and upholding the importance of due processes the SAI may be seen as an opponent of change. The SAI might be perceived as intrusive, creating distrust among the auditees.¹

At the same time, program evaluation is experiencing tremendous growth around the

world:

Right now, organizations such as the World Bank and the Organization for Economic Cooperation and Development work with governments at the senior level to create the recognition and demand for evaluation. These efforts are essential. . . . When meetings are organized, dozens, if not hundreds of people travel great distances to attend. Not just a few people. Not just an elite group. Not just dignitaries and officials. There is a tremendous demand for evaluators and users of evaluation to come together, and there is enormous power in these grass - roots networks to move the whole process forward. History is showing that if it is built, they will come.²

As this paper has demonstrated, those who attend and those who speak at evaluation association conferences and other meetings are frequently officials from state finance offices, audit offices, and controllers' offices, exchanging ideas with and learning from others in the evaluation network.

In researching what helps build an organization's capacity for program evaluation, GAO has found that

Demand for information on what works stimulated some agencies to develop an institutional commitment to evaluation. The agencies we reviewed did not appear to deliberately set out to build an evaluation culture. Rather, a systematic, reinforcing process of self - examination and improvement seemed to grow with the support and involvement of agency leadership and oversight bodies.³

GAO concluded also that

Whether evaluation activities were an intrinsic part of the agency's history or a response to new external forces, learning from evaluation allowed for continuous improvements in operations and programs, and the advancement of a knowledge base.⁴

For SAIs, the way seems to be clear: It is not possible to turn back.

[6.2] Pathways to New Learning

One of the mandates of INTOSAI's Working Group on Program Evaluation is to make information on evaluation available to all SAIs. This paper, and the resources it points to, partially fulfills that mandate. What does this mean for SAIs without adequate access to the Internet or for SAIs faced with traveling to conferences or seminars on slim budgets? It would appear to mean a redoubling of global support among the hundreds of evaluation professionals who are able to network, individually and in associations.

Many times each year, somewhere on the planet, a conference, seminar, or workshop in evaluation theory or methods is being conducted. News of these sessions is carried every day through the Internet, through the dozens of professional associations, through the burgeoning numbers of evaluators' Listservs, and by word of mouth even in the halls of SAIs on every continent. The work to be done is not just acquiring the learning that is being made available in these sessions; it is also learning how to make learning more widely accessible.

Many pathways are open to start with. New opportunities might be offered through INTOSAI's Development Initiative (IDI), for example, or its associated institutions. The American Evaluation Association holds conferences on evaluation theory, methods, and practice and is also allied with topical interest groups of international concern—including the group on International and Cross Cultural Evaluation, often chaired by members of The World Bank, Latin America's PREVAL, or like organizations. AEA's International and Cross Cultural Evaluation interest group has made available travel subsidies to professional evaluators who live and practice in developing countries so that they may attend its conferences.

It has already been noted that the number of associations, and therefore the number of evaluation professionals, is growing incrementally. Opportunity is all.

[6.3] Supporting an Evaluation Culture

The audit by SAIs of evaluation activities undertaken by the public administration may effectively help enhance a country's evaluation culture. Auditing may highlight the significance of evaluations for ensuring a responsible use of public funds, point out deficiencies in the use of this tool and generate recommendations for improvement. In Germany, where the administration is obliged under budgetary law to carry out programme results evaluations, the SAI (Bundesrechnungshof) published a report, as early as 1998, which received much support for carrying forward further evaluation efforts. As part of a horizontal audit mission, auditors found that the administration did not adequately comply with its obligation to conduct systematic and objective studies of programme effectiveness and efficiency. The Bundesrechnungshof considered these shortcomings to be the result of inadequate and ambiguous definition of policy programme objectives and lacking programming of subsequent evaluation work. Where the administration had carried out evaluation work, the findings were often not adequately used in decision-making on comparable or follow-up programmes. In line with the advice provided in the above report, federal government enhanced and specified the framework for evaluating public expenditure programmes. Currently, the Bundesrechnungshof is doing a cross-cutting study designed to measure the progress made and identify any problems that still persist in the public administration.

What, after all, is the goal of evaluation? On one hand, as members of the Netherlands Court of Audit pointed out at the 2002 conference of the European Evaluation Society, public auditing functions may be subject to a complex kind of insularity:

A government audit office can be considered as part of the institutionalised learning abilities of government. In the traditional policy cycle of preparing policies, implementing them, evaluating them and feeding back the results to adjust policies, the audit function is clearly positioned in the evaluative part of the cycle. Moreover, the relatively high status and visibility of a supreme audit institution increases the impact of its work. . . .

[G]overnment auditing can play a role in the learning behaviour of auditees,

i.e. government agencies. Auditing carried out by independent national audit offices is an institutionalised part of the policy cycle of most governments in the developed world. It involves the evaluation of financial accounts, financial management and—increasingly—of results of public policies and programs. This privileged position—implying a continuous relation with public auditees— helps to develop learning behaviour and ultimately could lead to improvement of public institutions and programs. Some warnings are appropriate, however. The independent position of most national audit institutions also implies relatively separated worlds of auditors on one hand and auditees on the other. Moreover, these institutions are in most cases monopolists by law In practice, this can translate into less relevant choices of audit topics and a blind eye for the many complicated features of policy making in the real world.⁵

On the other hand, to be truly effective, evaluation, as with all learning, needs to make a place for new ideas, new developments, and new people supporting one another. Paul Duignan, senior research fellow at the Centre for Social and Health Outcomes Research and Evaluation at New Zealand's Massey University, suggested at the 2001 American Evaluation Association meeting that

[W]hat's needed is that all people at all levels throughout our organizations and policy - making processes are being more evaluative about what they do. Note that this does not mean they all need to be calling what they do *evaluation*. They simply need to ensure that people become more *evaluative*. They must have appropriate evaluation skills, systems, structures, and resources to support them in taking a more evaluative approach to their work.

Looked at this way, the task of mainstreaming evaluation may be better put as one of *building evaluative capability*, *building evaluation capacity*, or even *building an evaluation culture* throughout our programs, organizations, and policy development. Such evaluative activity may not necessarily be labeled evaluation but it should contribute to the goal of being more evaluative and making organizations, policies, and programs more effective and efficient. . . . Looking at [mainstreaming] as the attempt to get people to be more evaluative, we may need to be prepared to give evaluation away in

order to build evaluation capability.⁶

From what base can old knowledge, or new, be given away? Arturo Gonzáles de Aragón, Auditor General of Mexico speaking in 2003 at the opening session of INTOSAI's 50th anniversary celebration had this to say:

INTOSAI has become a fundamental, strategic organization for the assessment, auditing, and development of public policies. Growing from the 34 countries that started on this road in the year 1953, to an organization that now includes 185 members, there have been a great number of accomplishments and renewed goals in which we all have taken part. . . .

The organization's committees and working groups, which aim to develop analyses, research, and guidelines on specific audit themes, are especially relevant for INTOSAI. Among them, we can point to the . . . Working Groups on Privatization and Environmental Auditing, [and] Program Evaluation

Audit institutions receive countless benefits from these efforts

...

. . . If the pluralistic world in which we live today is better than the world in 1953, when INTOSAI was formed, we can truly say that this organization has helped considerably toward that end. . . .

Thanks to INTOSAI, we have learned that our mission means devoting our main efforts to the welfare of the societies that have entrusted us with the great responsibility of serving others.⁷

Wrapping up the remarks at the anniversary ceremony, Franz Fiedler, Secretary General of INTOSAI, added in the same vein that after 1953 INTOSAI's

[D]ramatic increase in membership had an impact on the structures of INTOSAI and required a wider organisational basis, resulting in . . . the establishment of committees and working groups] [that, having been]

assigned specific themes for study, came to be the think tanks of INTOSAI. The basic documents that they elaborated on a range of subjects over the course of time are generally recognised as, and have turned out to be, practice - oriented guidelines for day - to - day audit work of INTOSAI members.

...

INTOSAI welcomes a new, modern - day strategy, as it has never been an organisation entrenched in outdated perceptions, but it must be a strategy undergoing permanent change. However, this change was never revolutionary, aiming at abrupt transformation, but evolutionary in nature, taking account of the interests of its members⁸

More recently, the International Organisation for Cooperation in Evaluation paralleled these remarks, noting that

We live in a world that demands continuous improvement in the performance of administrations; greater accountability and transparency; . . . and effective delivery of results in the public and private sectors and civil society. Evaluation addresses these demands by providing feedback on what has worked; deepening our understanding of the processes of policy implementation; designing knowledge systems that allow institutions and systems to learn; and developing capacities to manage effectively and to innovate. In many countries and in every continent evaluators have come together to form professional associations and societies. . . . In the 1980s there were only three national and regional evaluation societies, by the late 1990s there were nine and by the beginning of the 21st century the number had grown to fifty.⁹

Or, in the words of an old Chinese proverb, “Enough shovels of earth—a mountain. Enough pails of water—a river.”

- 1 Jane Etverk, “Measuring Performance Audit Effectiveness: The Case of Estonia,” M.A. thesis, Department of Public Administration, University of Tartu, Tartu, Estonia, June 27, 2002, pp. 16–17. National Audit Office of Estonia, About NAO, Documents, Other Documents, Tallinn, Estonia. www.riigikontroll.ee/?lang=en (Apr. 30, 2009).
- 2 A. J. Love and C. Russon, “Building a Worldwide Evaluation Community: Past, Present, and Future,” *Evaluation and Program Planning* 23:4 (2000): 450. (Apr. 30, 2009).
- 3 GAO, *Program Evaluation: An Evaluation Culture and Collaborative Partnerships Help Build Agency Capacity*, GAO - 03 - 454 (Washington, D.C.: May 2, 2003), p. 16. www.gao.gov (Apr. 30, 2009).
- 4 GAO, *Program Evaluation: An Evaluation Culture and Collaborative Partnerships Help Build Agency Capacity*, p. 24.
- 5 Gerard Bukkems and Hans de Groot, Netherlands Court of Audit, “Learning by Auditing: Mission Impossible?” paper for the 5th biennial conference of the European Evaluation Society, Three Movements in Contemporary Evaluation: Learning, Theory and Evidence, Seville, October 10–12, 2002, pp. 2 and 5. www.evaluationcanada.ca/distribution/20021010_bukkems_gerard_de_groot_hans.pdf. (Apr. 30, 2009).
- 6 Paul Dignan, “Mainstreaming Evaluation or Building Evaluation Capability? Three Key Elements,” in J. Jackson Barnette and James R. Sanders, eds., *The Mainstreaming of Evaluation*, New Directions for Evaluation No. 99 (San Francisco: Jossey - Bass, 2003), p. 12. Emphasis in original.
- 7 Arturo Gonzáles de Aragón, Auditor General of Mexico, “The Benefits of INTOSAI to Supreme Audit Institutions,” pp. 122–23 and 127, in International Organization of Supreme Audit Institutions, *INTOSAI: 50 Years (1953–2003)*, Special Publication (Vienna: 2004). INTOSAI: International Organization of Supreme Audit Institutions, Documents, INTOSAI, General www.intosai.org (Apr. 30, 2009).
- 8 Franz Fiedler, “50 Years of INTOSAI,” pp. 142 and 147, in International Organization of Supreme Audit Institutions, *INTOSAI: 50 Years (1953–2003)*.
- 9 International Organisation for Cooperation in Evaluation, “IOCE as Part of a Broader

Movement.” <http://ioce.net/overview/general.shtml> (Apr. 30, 2009).

Appendix A: Evaluators' Resources

The resources listed here will enable evaluators to increase their skills and develop their capacities for engaging in teamwork and with the evaluation community. They suggest opportunities for networking with other evaluators, obtaining fellowships and grants, studying in workshops and seminars or self - paced coursework, locating meetings and job opportunities around the world, and identifying related information. The list is not exhaustive. Most of the Web sites within these resources lead to many other resources.

1. Awards and prizes	Contact
<p>Canadian Evaluation Society www.evaluationcanada.ca/site.cgi?s=1&ss=0&_lang=en</p>	
<p>CES has an annual award for a Contribution to Evaluation in Canada, open to Canadian Evaluation Society members and nonmembers. Another award for Service to the Canadian Evaluation Society is open to members of the Canadian Evaluation Society. Click on Awards. Competitions in program evaluation are available in English and French to students in Canadian postsecondary institutions. Prizes include cash awards and the costs of attending and traveling to the Canadian Evaluation Society annual conference. Click on Student Competitions.</p>	<p>Canadian Evaluation Society Secretariat 1485 Laperriere Avenue Ottawa, Ontario Canada K1Z 7S8 Tel.: (613) 725.2526 Fax: (613) 729.6206 secretariat@evaluationcanada.ca</p>
<p>Italian Evaluation Association (Associazione Italiana di Valutazione) www.valutazioneitaliana.it</p>	
<p>AIV offers annual prizes for the best thesis on set topics in program evaluation. Click on Convegni e scuole, Iniziative extra AIV, Premi di Valutazione for application details.</p>	<p>Associazione Italiana di Valutazione Segretario nazionale Piazza Vittorio Emanuele 25 00185 ROMA c/o Parsec Tel.: 06.4465409 Fax: 06.45438049 segreteria@valutazioneitaliana.it</p>

2. Fellowships and coursework	Contact
<p>CCAF-FCVI www.ccaf-fcvi.com/english/index.html</p>	
<p>The International Assistance Program for Improved Governance and Accountability of this national nonprofit foundation helps strengthen comprehensive auditing in participating national audit offices. Instruction and practical audit experience are provided by the Office of the Auditor General of Canada and the Office of the Auditor General of Québec. CCAF - FCVI offers fellowships to professionals from selected national audit offices brought to Canada for 9 months of intensive on - the - job and classroom training.</p> <p>CCAF - FCVI also offers international symposia to provide opportunities to exchange experiences, perspectives, and plans and to compare strategies for advancing governments' effectiveness in accountability, governance, and performance reporting. Courses and programs have been given in Europe, the United States, Australia, South Africa, and Latin America. Click on International Business for details.</p>	<p>CCAF-FCVI Inc. 291 Olmstead Street Ottawa (Vanier), Ontario Canada K1L 7J9 Tel.: (613) 241-6713 Fax: (613) 241-6900 info@ccaf-fcvi.com</p>

3. Coursework online	Contact
<p>Centre for Program Evaluation www.edfac.unimelb.edu.au/cpe</p>	
<p>CPE, an evaluation and research center in the University of Melbourne's Graduate School of Education, provides developmental activities, such as specialized training courses, for clients outside the university. The "Course Information" link leads to descriptions of a postgraduate certificate in assessment and evaluation and a master of assessment and evaluation. A number of courses, as well as a forum, are offered online.</p>	<p>Centre for Program Evaluation Room 422, Alice Hoy Bldg. University of Melbourne Melbourne, Australia Tel.: +61 3 8344-8394 Fax: +61 3 8344-8490 cpe-enquiries@unimelb.edu.au</p>
4. Networking and job opportunities	Contact
<p>UK Evaluation Society www.evaluation.org.uk</p>	
<p>The Society provides "information about the activities of the UKES and news about other evaluation activities, jobs, contracts, events, resources in the UK and internationally." The Networks, Events, and Jobs/Tenders pages list job openings and networking opportunities, meetings, seminars, and workshops, many free, in the United Kingdom and Europe.</p> <p>Beginning evaluators may find the site useful, since "About us" states that UKES is a "member of the Academy of Social Sciences and has developed broad links with the international community—particularly the European Evaluation Society and International Organisation for Cooperation on Evaluation."</p>	<p>UK Evaluation Society 37 Star Street Ware, Hertfordshire England SG127AA Tel.: +44 (0) 1920 462411 Fax: +44 (0) 1920 462730 ukes@profbriefings.co.uk</p>

5. Networking and self - paced tutorials	Contact
<p>Nigerien Network of Monitoring and Evaluation www.pnud.ne/reuse/HOMEUK.HTML</p>	
<p>ReNSE is an informal group of monitoring and evaluation professionals whose aim is to foster capacity - building, setting up a forum that will help define standards, methodologies, and professional practices. The Self - Training in M&E link leads to several well - constructed, self - paced tutorials in French and English. The Strategic Alliances link leads to news of various evaluation activities, including funding activity. The Calls for Evaluators link opens to a page intended to facilitate free contact between African evaluators and other international agencies.</p>	<p>Mr. Robert Ndamobissi M&E Specialist UNICEF, BP 12481, Niamey (Niger) Tel.. (+227) 72 30 08 Fax (+227) 73 34 68 rnamobissi@unicef.org</p>
6. Networking and standards	Contact
<p>The Swiss Evaluation Society (SEVAL) www.seval.ch/en/index.cfm</p>	
<p>SEVAL fosters the exchange of information between evaluators in politics, administration, academia, nongovernmental organizations, and the private sector. It actively engages in improving experience in and the quality of evaluation and its diffusion. SEVAL's evaluation standards, define the demands on evaluation of all kinds and are directed toward all who participate in or influence evaluation.</p>	<p>SEVAL University of Fribourg Ressort Evaluation, Rectorat Av. de Beauregard 11 (1.217) CH-1700 Fribourg Tel.: +41 (0) 26 300 82 82 Fax: +41 (0) 26 300 96 55 secretariat@seval.ch</p>

7. Networking	Contact
<p>GOVTEVAL www.angelfire.com/id/govteval</p>	
<p>Based in Malaysia, GOVTEVAL is an unmoderated, global electronic discussion group open to anyone interested in public sector program evaluation. It is operated jointly with the Program for Public Sector Evaluation at the Royal Melbourne Institute of Technology, Australia. Among its objectives are to bring together practitioners and users of evaluation, academics, and researchers in public sector program evaluation from across the world to share experiences and views.</p>	<p>majordomo@nasionet.net</p>
8. Travel subsidies	Contact
<p>International & Cross Cultural Evaluation Topical Interest Group www.encompassworld.com/index.html</p>	
<p>I&CCE is affiliated with the American Evaluation Association. The group's purpose is to give evaluators interested in cross-cultural issues opportunities for professional development. For example, for 2005, I&CCE gave away \$7,500 in travel funds to AEA international grant recipients.</p>	<p>Tessie Tzavaras Catsambas EnCompass LLC 10500 Alloway Drive Potomac, MD 20854 USA Tel.: (301) 299-3266 Email: tcatsambas@encompassworld.com</p>

9. Evaluation - related data bank	Contact
<p>The Foundation Directory Online http://fconline.fdncenter.org</p>	
<p>A database of 96,000 foundations, corporate donors, and grantmaking public charities and their profiles, along with details of some 900,000 recent grants. Subscribing to The Foundation Center makes available grants, training opportunities, books and articles, and job postings, as well as materials on how to write grant proposals.</p>	<p>The Foundation Center 79 Fifth Avenue/16th Street New York, NY 10003-3076 USA Tel.: (212) 620-4230 Tel.: (800) 424-9836 Fax: (212) 807-3677 fdonline@foundationcenter.org</p>
10. Miscellaneous contacts	
<p>Many evaluation societies around the world offer discounts to members for the costs of attending and traveling to conferences, buying books from publishers on evaluation topics, and the like. Several post job opportunities.</p>	

Appendix B: References

This list does not exhaust the literature on program evaluation. It is a small selection of the major sources that are most widely used among practitioners of evaluation or that are generally regarded as indispensable in the field.

American Evaluation Association. *Guiding Principles for Evaluators*. Fairhaven, Mass.: July 2003. www.eval.org.

Bamberger, Michael J., Jim Rugh, and Linda Mabry. *RealWorld Evaluation: Working under Budget, Time, Data, and Political Constraints*. Thousand Oaks, Calif.: Sage Publications, 2006.

Chelmsky, Eleanor, and William Shadish. *Evaluation for the 21st Century: A Handbook*. Thousand Oaks, Calif.: Sage Publications, 1997.

Comptroller General of the United States. *Government Auditing Standards, July 2007 Revision*, GAO-07-731G. Washington, D.C.: U.S. Government Accountability Office, July 2007. www.gao.gov.

European Commission, Directorate – General for the Budget. *Evaluating EU Activities: A Practical Guide for the Commission Services*. Luxembourg: Office for Official Publications of the European Communities, July 2004. European Commission, Financial Programming and Budget Publications, http://ec.europa.eu/budget/library/publications/financial_pub/eval_activities_en.pdf.

Evaluation: The International Journal of Theory, Research, and Practice. Published by Sage Publications, in association with Tavistock Institute, London, England. www.uk.sagepub.com/home.nav. *Evaluation* can also be accessed through Sage

Journals Online at <http://online.sagepub.com>.

Evaluation and Program Planning: An International Journal. Published by Elsevier Ltd., Oxford, England.
www.elsevier.com/wps/find/journaldescription.cws_home/593/description#description.

Evaluation Checklists, The Evaluation Center, Western Michigan University, Kalamazoo, Michigan. www.wmich.edu/evalctr/checklists.

GAO. *Designing Evaluations*, GAO/PEMD - 10.1.4. Washington, D.C.: May 1991. www.gao.gov.

INTOSAI Auditing Standards Committee. *Implementation Guidelines for Performance Auditing: Standards and Guidelines for Performance Auditing Based on INTOSAI's Auditing Standards and Practical Experience*. Stockholm, Sweden: July 2004. INTOSAI: International Organization of Supreme Audit Institutions, Documents, INTOSAI, Audit - related, INCOSAI Goal 1. www.intosai.org.

Rossi, Peter H., Mark W. Lipsey, and Howard E. Freeman. *Evaluation: A Systematic Approach*, 7th ed. Thousand Oaks, Calif.: Sage Publications, 2003.

Sanders, James R., and The Joint Committee on Standards for Educational Evaluation. *The Program Evaluation Standards: How to Assess Evaluations of Educational Programs*. Thousand Oaks, Calif.: Sage Publications, 1994.

Shadish Jr., William R., Thomas D. Cook, and Laura C. Leviton. *Foundations of Program Evaluation: Theories of Practice*. Thousand Oaks, Calif.: Sage Publications, 1991.

Shaw, Ian F., Jennifer C. Greene, and Melvin M. Mark. *The Sage Handbook of Evaluation*. Thousand Oaks, Calif.: Sage Publications, 2006.

Stufflebeam, Daniel L., and Anthony J. Shinkfield. *Evaluation Theory, Models, and Applications*. Hoboken, N.J.: Jossey - Bass, 2007.

Appendix C: Glossary

This appendix defines terms related to program evaluation that are used in Program Evaluation for SAIs: A Primer. The definitions were taken from one or more of the following glossaries, four of which are available on the Internet at the addresses indicated.

1. European Commission, *Evaluating EU Activities*, European Commission, Directorate- General for the Budget, *Evaluating EU Activities: A Practical Guide for the Commission Services* (Luxembourg: Office for Official Publications of the European Communities, July 2004), Annex I, Glossary, pp. 101–10. European Commission, Financial Programming and Budgeting, Publications. http://ec.europa.eu/budget/index_en.htm (May 6, 2009). [EC]
2. The Evaluation Center, Western Michigan University, Kalamazoo, Michigan. Tools & Resources, Glossaries, Glossary of Program Evaluation Terminology. <http://ec.wmich.edu/glossary/prog-glossary.htf> (May 6, 2009). [WMich]
3. GAO, *Performance Measurement and Evaluation: Definitions and Relationships*, GAO-05-739 SP. Washington, D.C.: May 2005. www.gao.gov (May 6, 2009).
4. Sandra Mathison, ed., *Encyclopedia of Evaluation* (Thousand Oaks, Calif.: Sage Publications, 2005). [EE]
5. U.S. Department of Justice, Bureau of Justice Assistance, Center for Program Evaluation and Performance Measurement, Glossary. <http://www.ojp.usdoj.gov/BJA/evaluation/glossary/index.htm> (May 6, 2009). [BJA]

The definitions have been compressed and paraphrased from these sources. With few exceptions, terms defined in only one glossary have been omitted. The appendix also excludes terms related primarily to auditing, on the principle that Supreme Audit Institutions are familiar with such terms.

Many other glossaries defining terms used in program evaluation are available in paper versions and on the Internet. The five selected for adaptation here were chosen for their comprehensiveness and accuracy, as well as for being international in scope.

Abbreviations following each definition indicate which of the five sources they were adapted from: (1) EC, (2) WMich, (3) GAO, (4) EE, (5) BJA.

Term	Definition
Audit, Audit of an evaluation	An independent, objective, and systematic assessment of an organization's control, risk management, and governance processes or an independent, objective, and systematic assessment of the quality of an evaluation plan and its implementation and of the accuracy and validity of the evaluation's results and conclusions. Auditing standards may be applied to a government organization's achieving the purposes that programs and funds are authorized for and their economic and efficient operation. In recent years, audit and evaluation practices have been blended together as resources for decision makers responsible for public programs. [EC, WMich, EE, BJA]
Before-and-after design	A design that measures outcome indicators both before and after a treatment's implementation; any change may be attributed to the treatment, after alternative hypotheses have been ruled out. [BJA]
Case study, Case study design	A comprehensive description and analysis of a single project or program taken as a whole and in context; it may be based on an in - depth study of one or more specific cases or situations. [EC, WMich, BJA]
Comparison group	A group of individuals whose characteristics are similar to those of the participants in the program being evaluated (the experimental group). The individuals in the comparison group may receive different services, activities, or products from those in the group being evaluated or they may receive none at all; in any case, they do not receive the same treatment as the experimental group. In an evaluation, the two groups are compared to determine which of the program's services, activities, or products produced the expected change. [WMich, BJA]

Term	Definition
Constraint	Any limitation, such as available funds, in planning or implementing a program evaluation. [BJA]
Control group	A randomly assigned group for assessing the effect of a program's services, activities, or products on the program's participants. The same information is collected for people in the program (the experimental group) and people in the control group. The individuals in the control group have characteristics similar to those of the program participants but do not receive the services, activities, or products of the program being evaluated. In an evaluation, the two groups are compared to determine which of the program's services, activities, or products produced the expected change. [EC, BJA]
Cost-benefit analysis	An all - inclusive comparison of a program's social and private benefits with those of its related costs to determine whether the benefits exceed the costs and by how much. The aim of cost - benefit analysis is to select the alternative that maximizes the program's net benefits. [EC, EE, BJA]
Cost-effectiveness analysis	A judgmental comparison of a program's costs and benefits, in terms of time, effort, resources, or performance, in order to make comparisons with other ways to achieve the same objectives. Cost - effectiveness analysis focuses on the costs for outcomes restricted to a program's measurable goals. [EC, WMich, EE, BJA]
Cross-sectional data	Observations made about a program's participants at a single point in time. [BJA]

Term	Definition
Data analysis	The systematic application of statistical and logical techniques, including models, nonstatistical analysis and judgmental techniques, cost - benefit and cost - effectiveness analysis, and multicriteria analysis, to describe, summarize, and compare information about a program for use in an evaluation. [EC, BJA]
Delphi technique	A systematic method for obtaining group consensus; in each of several successive rounds, experts are told the views of the other experts in the previous round. The technique may use mailed questionnaires or some other method of controlled feedback to the respondents until consensus is reached. [EC, WMich, EE]
Design. See also Research design	The overall plan for conducting a program evaluation, describing such elements as questions to be addressed; units of analysis; comparisons to make; data collection and analysis plan; variables and concepts to measure; study boundaries in terms of population, geography, and time; level of precision needed to produce useful and credible results; management plan; and report schedule. [WMich, EE, BJA]
Effectiveness	The extent to which a program's stated goals or objectives are achieved in terms of both output and impact. Effectiveness requires a causal link between what is being evaluated and the intended change. [EC, EE, BJA]
Efficiency	A measure of how well a program has converted its inputs, or resources, such as funds, expertise, or time, to outputs, results, or effects, given its stated goals or objectives. [EC, EE, BJA]

Term	Definition
Evaluation, Evaluation project	A systematic assessment, consisting of an evaluative work plan within a logical framework, of the design, implementation, and results of a program or policy to determine its worth, relevance, efficiency, effectiveness, or sustainability. [EC, WMich, EE, BJA]
Evaluation team	The staff, consultants, and others who together plan, design, and conduct a program evaluation, developing data collection instruments, collecting and analyzing data, and writing the report. [BJA]
Evaluator	A person who uses credible and verifiable evidence to assess and form conclusions about a program's performance, planning, designing, conducting, and reporting a program evaluation. [WMich, EE]
Experimental design	A design for assessing the effects of a treatment or intervention. The evaluator assigns participants randomly to the study's treatment and control groups and manages participant selection; treatment type, order, and administration; how the data are recorded; and the performance of the statistical analysis. [WMich, EE, BJA]
Expert opinion	A method of collecting the considered judgments of experts in a particular field. [EC]
External evaluation	An evaluation conducted objectively by an evaluator outside the organization that is responsible for the program being evaluated. [EC, WMich, EE]
Field experiment, Field test	A randomized, controlled study of a program in a setting like the one where it is to be used, ranging from a preliminary investigation to a full - scale summative evaluation. [WMich, EE, BJA]

Term	Definition
Field notes	An evaluator’s written record of observations, interactions, conversations, situational details, and thoughts during a study period. [WMich, BJA]
Impact, Impact assessment, Impact evaluation	A program’s effect, positive or negative, foreseen or unforeseen, direct or indirect, intended or unintended; an impact evaluation may focus on a program’s broad, long - term results, a specific influence of the program on a desired outcome, or the program’s net impact. [EC, EE, BJA]
Implementation evaluation	An assessment of the extent to which a program is operating as intended. It typically assesses program activities’ conformance to statutory and regulatory requirements, program design, and professional standards or customer expectations. [GAO]
Internal validity. <i>See also Validity</i>	Confidence that an evaluation’s conclusions establish a causal link between the program and its effect; an evaluation design in which the causal link between the program’s intervention and the observed effects remains uncertain is a threat to internal validity. Other threats include unreliability of measures, attrition of participants, and statistical regression. [EC, EE, BJA]
Longitudinal study	A study of the same group over extended time. [EE, BJA]
Meta –analysis. <i>See also Meta – evaluation</i>	In the social sciences, a synthesis of multiple research studies of the same phenomenon to reach an overall conclusion; the studies may or may not be evaluative. [EE]

Term	Definition
Meta – evaluation	A systematic analysis of a set of existing evaluations of similar programs in order not just to summarize but to evaluate their quality, estimate overall program effects, develop support for hypotheses, or assess the evaluators' work. [WMich, EE, BJA]
Methodology	The procedure for collecting and analyzing information, such as case studies or Delphi technique. [BJA]
Monitoring. <i>See also</i> Participatory evaluation	An ongoing review of a program in which data are systematically collected on specified indicators to inform management and stakeholders about whether the program's activities meet specific standards or requirements. In participatory monitoring and evaluation, stakeholders decide together how to assess progress, collect and analyze data, and act on their findings; performance - based monitoring and evaluation combines traditional monitoring implementation with assessment of results. Program evaluations may use the results of monitoring information. [EC, EE, BJA]
Multivariate analysis	An analysis of the relationships between several variables. [BJA]
Norm	A single value or distribution of values that constitute a group's typical performance. [WMich]
Normative question	A question that requires an evaluator to compare what is happening in a program to the standard for what should be happening. [BJA]
Outcome	Enduring changes that result from a program's services or outputs, such as changes in a person's knowledge or skills. [EC, EE, BJA]

Term	Definition
Outcome evaluation	An assessment of the extent to which a program achieves its outcome - oriented objectives. It focuses on outputs and outcomes (including unintended effects) to judge program effectiveness but may also assess program process to understand how outcomes are produced. [GAO]
Output	Tangible measures of a program's results. For example, the number of persons a program trains are outputs. [EC, EE, BJA]
Panel data, Panel interviews	The collection of observations through repeated interviews with the same sample of respondents over a period of time. [BJA]
Participant. See also Stakeholder	An individual, family, neighborhood, or entity that receives a program's benefits, activities, or services. [EE, BJA]
Participatory evaluation	An evaluation designed, conducted, interpreted, and reported by the program's staff or oversight organization as well as its participants and other stakeholders, often with the aim of building evaluation capacity. [EE]
Performance audit	A systematic assessment of a program's efficiency, or of an intervention's direct outputs, and management; while performance audit and program evaluation both aim to improve quality, evaluation looks at the program's relevance and long - term consequences. [EC]
Performance evaluation	A comparison of a program's actual performance with what was planned in terms of resources and production, with the aim of improving the program by redirecting its resources and redesigning its structure. [BJA]

Term	Definition
Performance indicator. <i>See</i> Performance measure	
Performance measure, Performance measurement	An assessment of a program's progress toward, or achievement of, its goals or objectives, as observed against measurable indicators. Performance can be judged according to quality, efficiency, or effectiveness (both outputs and outcomes). [EE, BJA]
Policy	A governing principle that pertains to a set of activities that may be without limit in terms of time or money and that may benefit different people but that are directed toward common objectives whose issues are not resolved solely on the basis of fact or logic. [EC, BJA]
Process evaluation. <i>See</i> Implementation evaluation	
Qualitative data, Qualitative information	Facts and other information that may be presented as description or narrative representing categorization or interpretation. [WMich, BJA]
Quasi – experimental design	A design that has some characteristics of an experimental design; evaluators may use comparison groups and maximum controls to minimize threats to validity, but random selection is typically impossible or impractical. Alternatives are before - and - after comparisons, interrupted time - series, nonequivalent groups, and regression - discontinuity designs. [EC, EE, BJA]

Term	Definition
Reliability, Reliable measure	A measure that yields consistent, stable, and predictable results with repeated observations under similar conditions over time; the consistency or dependability of data with respect to the quality of the procedures and analyses evaluators used to collect and interpret them. [WMich, EE, BJA]
Replicate, Replication	The repetition or duplication of a program or an evaluation with all essentials unchanged. [WMich, BJA]
Research design	A plan for collecting the data for an evaluation that includes where and when to collect the data and how to analyze them. [EE, BJA]
Results	A program's effects, output, outcome, or impact, whether intended or unintended or positive or negative. [EC]
Results –based management	A management strategy that focuses on performing and achieving outputs, outcomes, and impacts. [EE]
Sample, Sample design	A set of individuals selected as a part or subset of a study population to be intentionally representative of that population; the procedure used to produce the sample. [EC, WMich, EE, BJA]
Sampling frame	A complete list of the population being studied in an evaluation, such as all the beneficiaries in the program. [BJA]
Scope, Scoping	An evaluation's field of investigation, including, among other things, the key objectives or issues the evaluation examines, a discussion of the validity of its issues, the time period and geographic area it covers, the assignment's complexity, the selection of the evaluation team, and an analysis of alternative ways for conducting the evaluation. [EC, BJA]

Term	Definition
Secondary data; Secondary data analysis	Data, often in large data sets, used in an evaluation that were originally collected for some other purpose; a reanalysis of such data, using the same or other appropriate procedures to verify the accuracy of the results of the initial analysis or to answer different questions. [WMich, EE, BJA]
Stakeholder. <i>See also</i> Participant	Individuals or groups that have a direct or indirect interest in a program and its evaluation. [EC, WMich, EE]
Terms of reference	An outline of work the evaluator is to do, including the evaluation's purpose, scope, and method; the questions to be answered; the resources and time allocated; and the reporting requirements. [EC]
Validity	The extent to which an instrument accurately measures what it is supposed to measure. Internal validity usually refers to a specific case; external validity refers to whether it is possible to generalize from a specific case to other populations, settings, or times. [WMich, EE, BJA]

Index of Studies

This index lists the program evaluations and performance audits that illustrate the concepts and methods in Program Evaluation for SAIs: A Primer, by page, type and title of document, and the SAIs that produced them. The index does not include documents produced by groups or persons who were not members of Supreme Audit Institutions.

Evaluation	Audit	Guidance	Document	SAI
		✓	Making Public Money Count: Wales Audit Office Strategy 2006–2009	Wales, Auditor General
		✓	Reporting on Outcomes: Setting Performance Expectations and Telling Performance Stories	Canada, Auditor General
		✓	Comparative Study of the Reports of Supervisory Boards: Examples and Tips for Better Reports	Netherlands, Court of Audit (Algemene Rekenkamer)
	✓		Public Service Management Reform: Progress, Setbacks and Challenges	Canada, Office of the Auditor General
	✓		Product Safety Monitoring	Finland, National Audit Office
✓			TCU Evaluation of the Actions for Detection and Correction of Visual Impairment	Brazil, Court of Audit (Tribunal de Contas da União)

Evalua tion	Audit	Guid ance	Document	SAI
		✓	Government-wide and Sectoral Performance Audit Reports	Philippines Commission on Audit
		✓	Organizational Culture: Techniques Companies Use to Perpetuate or Change Beliefs and Values	US. Government Accountability Office (GAO)
		✓	Strategy to Improve Planning for Performance Audits	Thailand, State Audit Office
		✓	Measuring Performance Audit Effectiveness: The Case of Estonia	Estonia, National Audit Office
		✓	TCU Annual Report, Fiscal Year 2002	Brazil, Court of Audit (Tribunal de Contas da União)
✓			Program Evaluation: An Evaluation Culture and Collaborative Partnerships Help Build Agency Capacity	US. Government Accountability Office (GAO)
		✓	Performance Auditing in the Australian National Audit Office	Australia, National Audit Office
		✓	Better Practice in Annual Performance Auditing	Australia, National Audit Office
		✓	Performance Audit	Netherlands, Court of Audit (Algemene Rekenkamer)
		✓	Government Auditing Standards, 2007 Revision	US. Government Accountability Office (GAO)

Evalua tion	Audit	Guid ance	Document	SAI
	✓		Anti-Erosion Programme in Rodrigues	Mauritius, National Audit Office
		✓	Evaluation Guidelines	Denmark, Ministry of Foreign Affairs
		✓	The Evaluation Synthesis	US. Government Accountability Office (GAO)
		✓	Performance Measurement and Evaluation: Definitions and Relationships	US. Government Accountability Office (GAO)
		✓	State Audit Law	Vietnam, Office of the State Audit
		✓	Performance Audit Manual	Bangladesh, Supreme Audit Institution
		✓	General Guidance on the Conduct of Performance Audits	Australia, Australia National Audit Office
		✓	Reporting Public Sector Performance, 2nd ed.	New Zealand, Office of the Auditor-General
✓			La Politique de la Ville	France, Cour des Comptes
		✓	Designing Evaluations	US. Government Accountability Office (GAO)

Evalua tion	Audit	Guid ance	Document	SAI
		✓	The Kingdom of Saudi Arabia's Message to the 7th Global Forum on Building Trust in Government for Further Participation and Transparency	Saudi Arabia, General Auditing Bureau
		✓	Key Success Factors for Effective Co- ordination and Collaboration between Public Sector Agencies	New Zealand, Office of the Auditor-General
✓			Co-ordination and Collaboration in the Criminal Justice Sector	New Zealand, Office of the Auditor-General
	✓		Performance Audit in Practice: Examples from County and Municipality Audits	Denmark; National Audit Office (Rigsrevisionen)