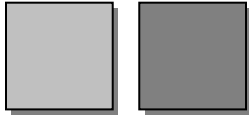IPDET

Handbook

Module 1

Introduction to Development Evaluation

Introduction

History shows that people have used evaluation in many ways and for many reasons for many centuries. Development is no exception. This first module of this course will introduce you to development evaluation. The module will serve as a guide to general information about development evaluation.

This module has six sections. They are:

- Evaluation, What Is It?
- Origin and History of the Evaluation Discipline
- The Evaluation Development Context
- Built on OECD DAC Criteria
- Principles and Standards
- Growth of Professional Evaluations Associations.
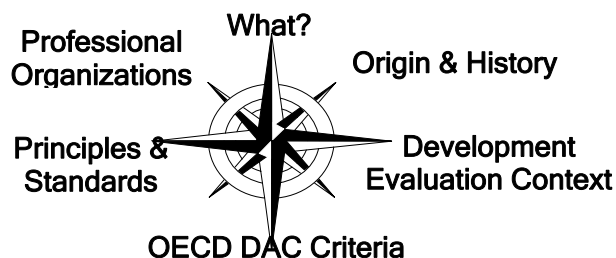
Learning Objectives

By the end of the module, you should be able to:

- define evaluation
- describe the purposes of evaluation
- discuss the history and origins of evaluation
- describe the changes in development evaluation
- describe the OECD DAC criteria
- describe the role and value of principles and standards in development evaluation.

Key Words

You will find the following key words or phrases in this module. Watch for these and make sure that you understand what they mean and how they are used in the course.

evaluation

formative evaluation

summative evaluation

prospective evaluation

monitoring

ex ante evaluation or prospective evaluation

internal evaluation

external evaluation

participatory evaluation

independence

audit

standard audit

financial audit

performance audit

relevance

effectiveness

efficiency

impact

sustainability

stakeholder analysis

rapid assessment

community mapping

evaluation synthesis

standards

utility

feasibility

propriety

accuracy

systematic inquiry

Evaluation: What Is It?

To begin understanding development evaluation, it is important that you understand what is meant by evaluation, its purposes, and how it can be used.

In this section you will learn about:

- defining evaluation
- purpose of evaluation
- using evaluation
- monitoring and evaluation.

Defining Evaluation

Evaluation can be defined in a number of ways. The straightforward dictionary definition of evaluation is:

> 1. the action of appraising or valuing (goods, etc); a calculation or statement of value; 2. the action of evaluating or determining the value of (a mathematical expression, a physical quantity, etc.), or of estimating the force of (probabilities, evidence, etc).[1]

Yet within the evaluation discipline, the term has come to have a variety of meanings. Indeed, there is no universal agreement on the definition itself.

> In fact, in considering the role of language in evaluation, Michael Scriven, one of the founders of modern evaluation, recently noted that there are nearly *sixty* [emphasis added] different terms for evaluation that apply to one context or another. These include: *adjudge, appraise, analyze, assess, critique, examine, grade, inspect, judge, rate, rank, review, score, study, test...*[2]

---

[1] *Oxford English Dictionary*, http://dictionary.oed.com/

[2] Jody Fitzpatrick, James Sanders, and Blaine Worthen. *Program evaluation: Alternative approaches and practical guidelines*, Third Edition (New York: Pearson Education, Inc. 2004), p. 5.

The standard Organisation for Economic Co-operation and Development (OECD) definition of evaluation, which we will use, is as follows:

OECD Evaluation Definition

Evaluation refers to the process of determining the worth or significance of an activity, policy or program.

An assessment, as systematic and objective as possible, of a planned, on-going, or completed intervention.

(See also "OECD Glossary of Key Terms in Evaluation and Results-Based Management," in your supplementary materials.)

In this context, it is important to introduce the notions of **formative evaluations,** and **summative evaluations** for projects, programs, and policies.

> Formative evaluations are evaluations intended to improve performance, [and] are most often conducted during the implementation phase of projects or programs. Formative evaluations may also be conducted for other reasons such as compliance, legal requirements or as part of a larger evaluation initiative. Summative evaluations, by contrast are studies conducted at the end of an intervention (or a phase of that intervention) to determine the extent to which anticipated outcomes were produced. Summative evaluation is intended to provide information about the worth of a program. (OECD Glossary).

A **formative evaluation** is an evaluation that looks into the ways in which the program, policy or project is implemented, whether or not the assumed 'operational logic' corresponds with the actual operations and what (immediate) consequences the implementation (stages) produce. This type of evaluation is conducted during the implementation phase of projects or programs. A part of what is known as "monitoring and evaluation" can also be seen as process (oriented) studies. Formative evaluations may include rapid appraisals, mid-term evaluations, and evaluations of implementation processes. Evaluations during the implementation phase (process evaluations) provide feedback so that the implementation can be improved and barriers to improved performance can be identified and removed.

Mid-point evaluations can begin to focus on lessons learned, as well as relevance, effectiveness and efficiency. Lessons learned are also important in guiding future interventions or changing current ones.

A **summative evaluation** *(outcome/impact evaluation)* is conducted at the end of an intervention (or a phase of that intervention) to determine the extent to which anticipated outcomes were produced and an impact was realized or not. Summative evaluation is intended to provide information about the worth and the impact of the program. Summative evaluations may include: impact assessments, cost-effectiveness investigations, quasi-experiments and randomized experiments.

Outcome or impact evaluations are generally conducted after the interventions are completed. They are used to answer questions of relevance, performance, impacts, sustainability, external utility and lessons learned.

Stated another way:

- formative evaluations focus on project/program/policy implementation and improvement

- summative evaluations focus on consequences or results — enabling persons to make assessments with respect to the creation, continuation or enlargement of a given project/program/policy.

A balance between the two kinds of evaluation is often needed. Frequently, *both* kinds of evaluation are needed and utilized by organizations – during the before cycle of a project, program, or policy.

Besides formative and summative evaluations, organizations can usefully employ **prospective evaluation**. Prospective evaluations are somewhat similar to evaluability assessments. They answer the questions "Is this program/project worth evaluating?" and "Will the gains be worth the effort/resources expended?" A prospective evaluation "synthesizes monitoring and evaluation information from earlier studies to assess the likely outcomes of proposed new projects [programs/policies]… [For example, US] Congressional committees and individual congressmen frequently ask the US General Accounting Office (GAO) for advice in forecasting the likely outcomes of proposed legislation…"[3] The GAO relies on its extensive database of socio-economic time-series data to inform such evaluations.

---

[3] Joseph Valadez and Michael Bamberger, editors. *Monitoring and evaluating social programs in developing countries: A handbook for policymakers, managers, and researchers.* (EDI Development Studies, World Bank, 1994), p. 63.

Such evaluations are performed less frequently in developing countries where such extensive time-series data may be absent; however, developing countries often do have monitoring and evaluation information which can be used to prepare a prospective evaluation of a project/program/policy.

A relatively old but nevertheless interesting example is the US GAO study 'Teenage pregnancy: 500,000 Births a Year but Few Tested Programs´ that was published in 1986. The GAO first outlined that the problem of teenage pregnancy indeed was severe and growing for unmarried mothers under 18 years old. It also found that only one federal demonstration program was focused exclusively on the problem, and that the evidence from demonstration programs, while sparse, suggested two types of future legislation. If expansion of services is essential, the first type of legislation would be flexible but targeted and would include both prevention and post-pregnancy services. The second would involve well-evaluated demonstrations of prevention and services that would be targeted, flexible, and innovative. To obtain this information, the GAO conducted an evaluation planning review in which it used four procedures. The GAO analyzed the main features of two congressional bills, reviewed available statistics on the extent of teenage pregnancy, examined the characteristics of federal and non-federal programs, and reviewed evaluation studies on the effectiveness of prior programs for assisting pregnant and parenting teenagers, as well as teenagers at risk of becoming pregnant. The GOA-evaluators reconstructed the underlying program theory and the operational logic of both congressional bills to find out why it was believed that these programs would work as proposed in the legislation. They then compared the evidence found to the features of the proposed legislation.

This type of evaluation is known as **ex ante** (before the fact) **evaluations or prospective evaluation**[4]. It can include program theory reconstruction / assessment and scenario studies as well as summarizing existing research to ascertain the empirical support for proposed initiatives.

---

[4] P. Rossi and H Freeman. (1993). *Evaluation: A Systematic Approach. Thousand Oaks:* Sage Publications. p. 422.

Purpose of Evaluation

Evaluation can be used for a variety of purposes. Again, within the discipline, there are different views about what the purpose or goal of evaluation should be within a given context.

A prevalent view that has emerged in the literature is that evaluation has four distinct purposes:

- an '**ethical**' purpose:
  to report to political leaders and citizens on how a policy has been implemented and the results achieved. This purpose combines the objectives of better accountability, information, and the serving of democracy

- a '**managerial**' purpose:
  to achieve a more rational distribution of financial and human resources among different actions, and to improve the management of the services entrusted with accomplishing them

- a '**decisional**' purpose:
  to pave the way for decisions on the continuation, termination, or reshaping of a policy

- an '**educative and motivational**' purpose:
  to help in educating and motivating public agents and their partners by enabling them to understand the processes in which they are engaged and identify themselves with their objectives.[5]

---

[5] French Council for Evaluation, Scientific and National Councils for Evaluation *A practical guide to program and policy evaluation* (National Council of Evaluation, 1999), p. 12

Prominent evaluators in the field describe the following purposes of evaluation:

- social betterment

- the fostering of deliberative democracy

- oversight and compliance

- accountability and transparency

- knowledge development and management

- organizational improvement

- promotion of dialogue and cooperation among key stakeholders

- project and/or policy relevance, implementation, efficiency, effectiveness, impact, and sustainability

- to generate lessons learned.

One expert takes a view highlighted in this IPDET course, that is, that the purpose of evaluation has become truly global. Chelimsky "…takes a global perspective, extending evaluation's context in the new century to worldwide challenges rather than domestic ones: new technologies, demographic imbalances across nations, environmental protection, sustainable development, terrorism, human rights, and other issues that extend beyond one program or even one country." [6]

Ultimately, the purpose of any evaluation is to meet the needs of users and stakeholders, whoever they may be.

---

[6] E. Chelimsky, "The coming transformations in evaluation," in E. Chelimsky and W.R. Shadish (eds), *Evaluation for the 21st Century: A handbook.* (Thousand Oaks, CA: Sage), 1997, and *Program Evaluation: Alternative Approaches and Practical Guidelines,* Jody Fitzpatrick, James Sanders, Blaine Worthen, Third Edition, (New York: Pearson Education, Inc. 2004), p. 11.

Who Benefits from Evaluation?

As you learned, evaluation has many benefits. But who actually benefits from evaluation. The following is a list of some of the people who can benefit from evaluation.

- government officials/parliament
- program managers and staff
- citizens
- businesses
- NGOs
- civil society
- donors
- participants.

What to Evaluate?

Evaluations can look at many different facets of development. The following are some of these.

- **Projects**: a single intervention in one location or a single project implemented in several locations.

- **Programs**: an intervention comprising various activities or projects, which are intended to contribute to a common goal.

- **Policies**: evaluations of the standards, guidelines or rules set up by an organization to regulate development decisions

- **Organizations**: multiple intervention programs delivered by an organization.

- **Sectors**: evaluations of interventions across a specific policy arena, such as education, forestry, agriculture, and health.

- **Thematic**: evaluations of particular issues, often cross-cutting, such as gender equity, global product goods, or millennium goal development.

- **Country assistance**: evaluations of progress relative to the plan, the overall effect of aid, and lessons learned.

Table 1.1 gives examples of evaluations.

Table 1.1: Examples of Evaluations

	Privatizing Water Systems	Resettlement
Policy Evaluations	Comparing model approaches to privatizing public water supplies	Comparing strategies used for resettlement of villagers to new areas
Program Evaluations	Assessing fiscal management of government systems	Assessing the degree to which resettled village farmers maintain previous livelihood
Project Evaluations	Assessing the improvement in water fee collection rates in two provinces	Assessing the farming practices of resettled farmers in one province

Uses of Evaluation

The results of evaluation can be used in many ways. Evaluations give clients, government agencies, NGOs, the public, and others feedback on policies, programs, and projects. The results can give information about how public funds are being used. Results can give managers and policy-makers information on what is working and what is not working, according to original objectives.

Evaluations can help to make policies, programs, and projects accountable for how they use public funds. They can also help stakeholders learn more about their policies, programs, and projects, leading to improvement of these.

Funding and ownership can also benefits from the results of evaluation. Carol Weiss[7] stresses the importance of identify the intended uses for an evaluation from the initial planning stage.

> … if you cannot identify and articulate the primary intended users and uses of the evaluation you should not conduct the evaluation. Unused evaluation is a waste of precious human and financial resources.

Weiss stresses that from beginning to end, the evaluation process is designed and carried out around the needs of the primary intended user. These primary users will have the responsibility for implementing change based on their involvement in the process or with the evaluation findings.

---

[7] Carol Weiss (2004). Identifying the intended use(s) of an evaluation. The IDRC. http://www.idrc.ca/ev_en.php?ID=58213_201&ID2=DO_TOPIC p 1

She also stresses the importance of negotiating between the evaluator or evaluation team and the primary intended users to determine ways to use the evaluation. Most evolutions will have many uses. By communicating among the evaluators and the intended users, you can establish a consensus about the uses.

Evaluation can:

- help analyze why intended results were or were not achieved

- explore why there may have been unintended results or consequences

- assess how and why results were affected by specific activities

- shed light on implementation processes, failures, or successes that may occur at any level

- help to provide lessons, highlights areas of accomplishment and potential, and offers specific recommendations for improvement and reform.

Evaluation professionals have developed a number of both pragmatic and more theoretical/conceptual ways of looking at how evaluation is or can be utilized. We will focus here on the pragmatic uses of evaluation.

Pragmatic Uses of Evaluation

- Help make resource allocation decisions.

- Help rethink the causes of a problem

- Identify emerging problems.

- Support decision-making on competing or best alternatives.

- Support public sector reform and innovation.

- Build consensus on the causes of a problem and how to respond.

Source: Kusak and Rist, p. 115.

As the following box summarizes, evaluation can help provide information on strategy, operations, and learning.

Evaluation provides information on:

Strategy: **Are the right things being done?**

      rational or justification

      clear theory of change

Operations: Are things being done right?

      effectiveness in achieving expected outcomes

      efficiency in optimizing resources

      client satisfaction

Learning: Are there better ways?

      alternatives

      best practices

      lessons learned

Source: Kusak and Rist, p. 117

In summary, managers can use information from evaluation to focus on:

- the broad political strategy and design issues ("Are we doing the right things?")

- operational and implementation issues ("Are we doing things right?")

- whether there are better ways of approaching the problem ("What are we learning?").

Monitoring and Evaluation

According to the OECD Glossary[8] "**Monitoring** is a continuing function that uses systematic collection of data on specified indicators to provide management and the main stakeholders of an ongoing development intervention with indications of the extent of progress and achievement of objectives and progress in the use of allocated funds."

In other words, monitoring is a routine, ongoing, and internal activity. It is used to collect information on a program's activities, outputs, and outcomes in order to measure the performance of the program.

Both monitoring and evaluation measure and assess performance, but they do this in different ways and at different times.

- Monitoring takes place throughout the project implementation

- Evaluation is the periodic assessment of the performance of the project.

Monitoring is an internal activity carried out by project staff and is generally the project management's responsibility to see that it happens and that the results are used. On the other hand, evaluation can be carried out internally or externally and it is the responsibility of the evaluator together with programme staff members.[9] Table 1.2 compares monitoring and evaluation.

Table 1.2 Comparison of Monitoring and Evaluation.

Monitoring	Evaluation
Ongoing, continuous	Period and time bound
Internal activity	Can be internal, external, or participatory
Responsibility of project staff and management	Responsibility of evaluator together with staff and management
Continuous feedback to improve program performance	Periodic feedback

---

[8] DAC (2002). *OECD Glossary of Key Terms in Evaluation and Results Based Management.* pp. 27-28. Also online at: http://www.oecd.org/dataoecd/29/21/2754804.pdf

[9] M&E In's and Out's. *InsideOut*, Issue #3: October/November 2005. Online at: http://www.insideoutresearch.co.za/docs/newsletter3.pdf

Roles and Activities of Professional Evaluators

As the concept and purposes of evaluation have evolved over time, so too have the roles and activities of evaluators. Professional evaluators have played a multitude of roles and engaged in numerous activities. Their role depends on the nature and purpose of the evaluation.

> Evaluators play many roles including scientific expert, facilitator, planner [as well as judge, trusted person, teacher, social change agent], collaborator, aid to decision makers and critical friend.[10]

Who Conducts the Evaluation?

Evaluators may be part of internal, external, or participatory evaluations.

Internal, External, and Participatory Evaluation

The OECD Glossary defines **internal evaluation** as "Evaluation of a development intervention conducted by a unit or individuals reporting to the management of the donor, partner, or implementing organization.[11]"

The OECD Glossary defines **external evaluation** as "Evaluation of a development intervention conducted by entities and/or individuals outside the donor, partner and implementing organization.[12]"

The OECD Glossary defines participatory evaluation as "Evaluation method in which representatives of agencies and stakeholders (including beneficiaries) work together in designing, carrying out and interpreting an evaluation".[13]

---

[10] Fitzpatrick, Sanders, and Worthen, *Program evaluation: Alternative approaches and practical guidelines*, p. 28.

[11] OECD (2002). *Glossary of key terms in evaluation and results based management*. Paris: OECD Publications. p. 26.

[12] OECD, *Glossary of key terms in evaluation and results based management.* p 23.

[13] OECD, *Glossary of key terms in evaluation and results based management.* p 28.

Advantages and Challenges of Internal and External Evaluators

There are advantages and disadvantages to using internal and external evaluators.

Internal evaluators usually know more about the program, project, or policy than an outsider. They usually know more about the history, organization, culture, people involved, and the problems and successes. Because of this knowledge, internal evaluators might be able to ask the most relevant and pertinent questions, know where to go "backstage" in the organization to find out what "really" is going on. This advantage, however, can also be a disadvantage. Internal evaluators may be so close to the program, project, or policy that they do not see it clearly and might not be able to recognize solutions or changes that others might see. Internal evaluators may also have the disadvantage of being more subject to pressures or influence from program decision-makers who also make personnel decisions. They may see the whole organization from only their own position inside the organization. Lastly, external stakeholders may perceive their findings as less credible.

External evaluators usually have more credibility and give the perception of objectivity to an evaluation. In addition, most external evaluators have more specialized skills that may be needed to perform effective evaluations. They also are independent from the administration and financial decisions about the program.[14] However, an external evaluation is not a guarantee of independent and credible results, particularly if the consultants have prior program ties. Just because a unit external to the organization conducts an evaluation does not guarantee it is independent and credible. External consultants, for example, may have prior program ties.

---

[14] Fitzpatrick, Sanders, and Worthen (2004). *Program evaluation: Alternative approaches and practical guidelines, 3rd edition.* Boston: Allyn and Bacon (Pearson). pp. 23 – 24.

Participatory Evaluation

Participatory evaluation is an "evaluation method in which representatives of agencies and stakeholders (including beneficiaries) work together in designing, carrying out and interpreting an evaluation." (OECD Glossary definitions). Participatory evaluation differs from internal and external evaluation in some fundamental ways.

> Participatory evaluation represents a further and more radical step away from the model of independent evaluation… [it] is a form of evaluation where the distinction between experts and layperson, researcher and researched, is deemphasized and redefined…evaluators… [act as] mainly facilitators and instructors helping others to make the assessment.[15]

Note the distinction between participatory evaluation and participatory methods; the latter may be used in internal and external evaluations.

Recently development evaluation has had a greater emphasis on participatory evaluations.

Evaluator Activities

Evaluators carry out a number of activities to match various roles. For example, internal evaluators may work on design, implementation, and outreach strategies for the intervention to be assessed. External evaluators would typically limit their involvement in program management. Regardless, generally, evaluators:

- consult with all stakeholders

- manage evaluation budgets

- perform or conduct the evaluation or hire and contract staff to perform the evaluation

- identify standards for effectiveness (relying on authorizing documents or other sources)

- collect, analyze, interpret, and report on data and findings.

As such, evaluators are expected to have a diverse skill set. In this context, evaluators can also play the role of knowledge creator/builder and disseminator of lessons learned.

---

[15] Stefan Molund and Göran Schill (2004). *Looking back, moving forward: SIDA evaluation manual.* Stockholm: SIDA. p. 19.

The Origins and History of the Evaluation Discipline

The modern discipline of evaluation emerged from social science research and the scientific method. However, evaluation has ancient traditions. Indeed, the earliest forms of evaluation thousands of years ago. For example, archaeological evidence shows that the ancient Egyptians regularly monitored their country's outputs in grain and livestock production more than 5000 years ago. "In the public sector, formal evaluation was evident as early as 2000 BC, when Chinese officials conducted civil service examinations to measure the proficiency of applicants for government positions. And, in education, Socrates used verbally mediated evaluations as part of the learning process."[16]

Some experts trace the emergence of modern evaluation methods to the advent of the natural sciences and attendant emphasis on observed phenomena (the "empirical method") in the 17th century.

> In Sweden, ad hoc policy commissions that performed some kind of evaluations came into being in the 17th century. "Traditionally, appointed ad hoc policy commissions have played a great part in preparing the ground for many decisions. The commissions have been important for providing briefing and (evaluative) background materials both with respect to fundamental policy decisions and in connection with the day-to-day fine-calibration of the arsenal of means available to various spheres of activity."[17]

Indeed, the commission system is still used in Sweden today, with several hundred currently in existence.

---

[16] Fitzpatrick Sanders, and Worthen (2004). *Program evaluation: Alternative approaches and practical guidelines, 3rd edition.* Boston: Allyn and Bacon (Pearson). p 31.

[17] J.-E. Furubo, and R Sandahl. (2002) "Coordinated Pluralism," *International atlas of evaluation,* J.-E Furubo, Ray C. Rist, and R Sandahl, editors. New Brunswick, NJ: Transaction Publishers. p. 116.

In the 1800's, evaluation of education and social programs began to take root in several Anglo-Saxon countries. Program evaluation was conducted in England by government-appointed commissions who were called upon to investigate and evaluate dissatisfaction with educational and social programs. The current-day external inspectorates for schools grew out of these earlier commissions. In the United States, pioneering efforts were made to examine the quality of the school system using achievement tests. These efforts have also continued to the present day, where student achievement scores are used to determine the quality of education and schools.

During this period, the early beginnings of accreditation for secondary schools and universities were to be found in the United States

In the early 1900's, formal evaluation and accreditation of medical schools took place in the United States and Canada. Other areas of investigation/measurement and evaluation during this period included: health, housing, work productivity, democratic and authoritarian leadership, and standardized educational testing. Most were small-scale efforts conducted by government agencies and social services.

In the development arena, "Dodd's attempt to introduce water boiling as a public health practice in villages in the Middle East is one of the landmark studies in the pre-World War II empirical sociological literature."[18]

Applied social research grew rapidly following the Great Depression when the United States' President Roosevelt instituted the New Deal socio-economic programs. The United States' federal government began to grow rapidly, as new agencies were created to manage and implement national programs for: agricultural subsidies to protect farmers; public works/job creation schemes; rural electrification; social security administration; the Tennessee Valley Authority, etc. As many of these large-scale programs were new and experimental in nature, the need for evaluating the effectiveness of them in jump-starting the economy, creating jobs, and instituting social safety nets grew in tandem.

---

[18] P. Rossi. and H. Freeman (1993). *Evaluation: A systematic approach.* Thousand Oaks, CA: Sage Publications. p. 10.

The need for evaluation increased during and after World War II, as more large-scale programs were designed and undertaken for the military, urban housing, job and occupational training, and health. "It was also during this time that major commitments were made to international programs for family planning, health and nutrition, and rural community development. Expenditures were very large and consequently were accompanied by demands for 'knowledge of results."

The Soviet launch of the first artificial satellite, Sputnik, in 1957 set off a superpower space race. The launch prompted American leaders to look at the American education system, and specifically at ways in which the teaching and achievement levels in math and science could be dramatically improved. The United States' government enacted the National Aeronautics and Space Act in 1958, which resulted in the development of new educational curricula and the creation of NASA.[19] Evaluation of these efforts was undertaken, albeit with less than fully satisfactory results because of problems inherent in social science evaluation design and methods.

In the 1950's and 1960's, evaluation became more routinely used in the United States and Europe to assess programs related to education, health, human services, mental health, prevention of delinquency, and rehabilitation of criminals. In addition, United States' President Johnson's "war on poverty" programs during the 1960's stimulated increased interest in evaluation. Work in developing countries around the world also expanded, with some evaluation activity for programs in agriculture, community development, family planning, health care, and nutrition. Again, for the most part, these assessments relied on traditional social science tools, such as surveys and statistical analysis.

[19] National Aeronautics and Space Act of 1958, http://www.hq.nasa.gov/office/pao/History/spaceact.html

In the late 1960's, many West European countries began to undertake program evaluation as well. In the Federal Republic of Germany, for example, Parliament started to require the federal government to report on implementation and impact of various socio-economic and tax programs. These included reports on the:

> joint federal-state program to improve the regional economic structure (1970); labor market and employment act (1969)…; hospital investment program (1971)…; general educational grants law (1969)…; and various reports on subsidies and taxes (1970 to present).[20]

During this time, the Canadian government also began to undertake steps toward evaluating government programs and performance. Canadian government departments were encouraged to establish planning and evaluation units. However, early efforts did not yield significant results. In the Canadian, German and Swedish cases, "…despite institutionalization of program evaluation in various policy areas, their systems remained rather fragmented and the number of studies carried out seems to be relatively low."[21]

In the late 1960's and early 1970's, numerous articles and books on the topic of evaluation began to appear in the United States and some OECD countries. Graduate school university programs[22] focusing on evaluation were developed to train a new cadre of evaluators to meet the increasing demands for accountability and effectiveness in government-financed socio-economic programs, such as elementary and secondary education grants, and "Great Society" (poverty reduction, Head Start preschools, civil rights, job corps, food stamps, etc.) programs. The Elementary and Secondary Education Act (ESEA) of 1965 mandated the government to assess student performance and teacher quality standards and provided resources (first US government budgetary set-aside for evaluation) to undertake these activities, thereby institutionalizing evaluation.

---

[20] Hans-Ulrich Derlien (1999). "Program evaluation in the Federal Republic of Germany," in *Program evaluation and the management of government: Patterns and prospects across eight nations* (Editor, Ray C. Rist). New Brunswick NJ: Transaction Publishers. p. 41.

[21] Hans-Ulrich Derlien (1999). "Genesis and structure of evaluation Efforts in comparative perspective," in *Program evaluation and the management of government: Patterns and prospects across eight nations (*Editor, Ray C. Rist). New Brunswick, NJ: Transaction Publishers. p. 148.

[22] See Annexes for list of evaluation training programs available around the world.

Canada, Sweden, and the Federal Republic of Germany (FRG) undertook program evaluation in the 1960's to assess new government-financed education, health, and social welfare programs.

> In this context formal planning systems emerged, which either were limited to medium-term financing planning (in the FRG) or even attempted to integrate budgeting with programming (in Sweden and Canada). In any case, evaluation was either regarded logically as part of these planning systems or as necessitated by the information needs of the intervention programs... Evaluations, then, were primarily used by program managers to effectuate existing and new programs.[23]

From the mid-1970's to the mid-1980's, evaluation became a full-fledged profession in many OECD countries. Professional evaluation associations were created, more programs to train evaluators were introduced, evaluation journals proliferated, and evaluation began to expand beyond the purview of government-financed programs to corporations, foundations, and religious institutions. For example, in France, public policy evaluation has been more systematically developed, with many universities—including the Grandes Ecoles— offering courses and /or information about evaluation as part of their curriculum.

The evaluation literature has also grown in quantity and quality.[24] Many OECD countries have established evaluation-training programs for civil servants either within the government or with outside contractors.

In addition, methodologies and models were explored, with greater emphasis on the information needs of consumers, examination of unintended outcomes, and the development of values and standards. Since 1985, computers and technology have vastly increased the ability of evaluators to collect, analyze and report on evaluation findings.

---

[23] H.-U. Derlien, (1999). "Genesis and structure of evaluation Efforts in comparative perspective," in *Program evaluation and the management of government: Patterns and prospects across eight nations.* Ray C. Rist, editor, New Brunswick: N.J.: Transaction Publishers. p. 153-154.

[24] C. Fontaineand E. Monnier (2002). "Evaluation in France," in *International atlas of evaluation,* Furubo, J.-E., Rist, Ray C., Sandahl, R. editors. New Brunswick, NJ: Transaction Publishers. p. 71.

Most evaluators recognize that evaluation is a complex political and technical effort. As we have seen, public policy, management, and administration have been the primary driving forces for the greater use of evaluation and development of the field. Indeed,

> [e]valuation research is more than the application of methods. It is also a political and managerial activity, an input into the complex mosaic from which emerge policy decisions and allocations for the planning, design, implementation, and continuance of programs to better the human condition. In this sense, evaluation research also needs to be seen as an integral part of the social policy and public administration movements.[25]

The need for results-based monitoring and evaluation has become a growing global phenomenon, as governments and organizations from around the world grapple with growing internal and external pressures and demands for improvement and reform in public management. The issue of accountable government performance has become a global phenomenon, as demands for transparency and accountability, demonstrable results, and program effectiveness have increased. This has led to the development of the results-based monitoring and evaluation model (see Module 4: *Building a Results-Based Monitoring and Evaluation System*), a powerful public management tool that be used to help leaders/managers track progress and demonstrate the impact of a given project, program or policy.

---

[25] Rossi, and Freeman. (1993). *Evaluation: A systematic approach.* Thousand Oaks: Sage Publications. p. 15.

Origins of Development Evaluation

Development evaluation evolved out of the audit and social science traditions. There are important similarities and differences, as well as linkages between these two traditions.

Audit Tradition

Auditing traces its roots to 19th century Britain when growing commercial and industrial development gave rise to the

> …need for verifiably accurate and dependable financial records… Auditors' work lent credibility to the growing capitalist infrastructure of the West. Auditors' opinions carried weight because of their technical craftsmanship and because auditors were outsiders.[26]

The **audit tradition** has an investigative, financial management, and accounting orientation: did the program do what was planned and was the money spent within the rules, regulations, and requirements? It uses concepts such as internal controls, good management/governance, and verification. Its emphasis is on accountability and compliance. Auditors are traditionally independent from program managers.

To take one example, France has a strong tradition of auditing, where the State Audit Office

> is responsible for judging the appropriate use of public funds. It evaluates financial data in order to identify the use of allocated funds, conducts case studies of administrative procedures, and interviews the relevant authorities.[27]

---

[26] R.A. Brooks (1996). "Blending two cultures: State Legislative auditing and evaluation," in *Evaluation and auditing: Prospects for convergence,* Carl Wisler, editor. *New Directions for Evaluation,* Number 71, Fall 1996. San Francisco: Jossey-Bass Publishers. p. 16,

[27] Fontaine and Monnier (2002). "Evaluation in France", in *International atlas of evaluation,* Furubo, J.-E., Rist, R.C., Sandahl, R. editors. New Brunswick, NJ: Transaction Publishers. p. 68.

Inspectors are assigned to ministries to audit and control "the legality of procedures and accounts."

There are several different types of auditing:

- **standard audit**: an independent, objective assurance activity designed to add value and improve an organization's operations. It helps an organization accomplish its objectives by bringing a systematic, disciplined approach to assess and improve the effectiveness of risk management, control and governance processes. (OECD glossary)

- **financial audit**: an audit that focuses on compliance with applicable statutes and regulations (OECD glossary)

- **performance audit**: an audit that is concerned with relevance, economy, efficiency, and effectiveness. (OECD glossary).

The auditing profession (unlike many evaluation sub-disciplines) has a common set of standards by which auditors abide. Indeed, auditing

> …gets much of its strength from the fact that it has a largely agreed upon set of standards (Institute of Internal Auditors and national standards). It delivers a range of products, from comprehensive to compliance audits, dealing with different aspects of an organization, and moves outward to the organization's activities and products.[28]

Furthermore, internal auditing "…encompass[es] [a wide array of] financial activities and operations including systems, production, engineering, marketing, and human resources."[29] It should also be noted that auditing gains strength too from the fact that professional accreditation is offered, which is not yet the case with evaluation.

---

[28] "Linkages Between Audit and Evaluation in Canadian Federal Developments," http://www.tbs-sct.gc.ca/pubs_pol/dcgpubs/TB_h4/evaluation03_e.asp

[29] The Institute of Internal Auditors," http://www.theiia.org

Since the 1940s, more and more developed and developing country governments created internal audit agencies or supreme audit institutions (SAI's) to examine public revenues and spending, and ensure financial accountability. Since the 1960s, new regional SAI's have also been established, including the Organization of Latin American and Caribbean Supreme Audit Institutions (1965), the Asian Organization of Supreme Audit Institutions (1978), and the European Organization of Supreme Audit Institutions (1990).

The International Organization of Supreme Audit Institutions (INTOSAI), established in 1953 with 34 member countries (and now has more than 170), is the professional organization of national SAI's. It "...issues international guidelines for financial management and other areas, develops related methodologies, provides training, and promotes the exchange of information among members."[30] Members adopted the Lima Declaration of Guidelines on Auditing Precepts, which called for independent government auditing.

Audit and Evaluation

Auditing and evaluation can be viewed as a kind of continuum, providing related but different kinds of information about compliance, accountability, impact, and results. There is some

> ...overlap in areas such as efficiency of operations and cost effectiveness...with evaluation concerned with analysis of policy and outputs, and auditing with internal financial controls and management systems.[31]

Auditing and evaluation have common objectives in that both aim to help decision-makers "...by providing them with systematic and credible information that can be useful in the creation, management, oversight, change, and occasionally abolishment of programs."[32]

---

[30] The International Organization of Supreme Audit Institutions," http://www.gao.gov/cghome/parwi/img4.html

[31] "Linkages between audit and evaluation in Canadian Federal developments," http://www.tbs-sct.gc.ca/pubs_pol/dcgpubs/TB_h4/evaluation03_e.asp

[32] C. Wisler, editor (1996) *Evaluation and auditing: Prospects for convergences*, in *New Directions for Evaluation*, Number 71, Fall 1996. San Francisco: Jossey-Bass Publishers. p.1.

Auditing and evaluating share other elements in common. For example, independence – critical to auditors – is equally important to evaluators. "Independent and self-evaluation are to the public sector what accounting and auditing are to the private sector."[33]

Much has been written on the differences between auditing and evaluation.[34] These differences stem from their origins, with auditing deriving much of its approach from financial accounting, and evaluation deriving much of its approach from the social sciences. One difference is that auditing focuses on *outputs*, while evaluation focuses on *outcomes*. Another key difference is the "inclination of auditors toward normative questions [what is versus what should be] and [the inclination] of evaluators toward descriptive and impact [cause and effect] questions."[35] On balance, it should be noted that social science has become more influential in evaluation than the audit traditions.

Social Science Tradition

As explained earlier, evaluation grew out of the social science tradition. As governments and organizations moved from questions about verification and compliance to questions about impact and results, social science techniques were incorporated into evaluation. Social science research methods and applied social research were first used by many OECD member country governments to assess education, health, and community programs.

Evaluation relies on both scientific and social research methods. The scientific method "refers to research methodologies that pursue verifiable knowledge through the analysis of empirical data."[36] The scientific method, as the name implies, grew out of the 13th century natural sciences, and relied on Aristotle's notion of induction.

---

[33] R. Picciotto, "Development Evaluation as a Discipline," p. 7.

[34] See example, *Evaluation and auditing: Prospects for convergences,* Carl Wisler, editor*, New directions for evaluation,* San Francisco: Jossey-Bass Publishers, Number 71, Fall 1996.

[35] See example, *Evaluation and Auditing: Prospects for Convergences,* Carl Wisler, editor, *New Directions for Evaluation,* Jossey-Bass Publishers, Number 71, Fall 1996, p. 3.

[36] *Dictionary of the social sciences.* Calhoun, C., editor. Oxford University Press, 2002.

The scientific method contains four basic steps:

1. Observation/Characterization/Identification of a problem

2. Formation/Development of a hypothesis

3. Predictions based on hypothesis

4. Experimentation/Testing of the predictions through data collection and analysis.

These steps can be repeated until an appropriate explanation for the problem can be derived. The advantage of the scientific method is that it can provide unbiased, scientific, and replicable/verifiable evidence.

Evaluation also uses a variety of methods from the social sciences, including sociology, anthropology, statistics, political science, etc. Indeed,

> The application of social research methods to evaluation coincides with the growth and refinement of the methods themselves, as well as with ideological, political, and demographic changes that have occurred this [past] century. Of key importance were the emergence and increased standing of the social sciences in universities and increased support for social research. Social science departments in universities became centers of early work in program evaluation and have continued to occupy an influence place in the field."[37]

Evaluation also draws heavily from social science research, in areas such as: theory construction, design, approach, data collection methodology, analysis and interpretation, statistics, surveys, and sampling.

Indeed, the origins and basis of many of the IPDET modules presented here will be drawn from the social sciences.

---

[37] Rossi and Freeman (1993). *Evaluation: A systematic approach.* Sage Publications, CA. p. 9.

The Evaluation Development Context

Evaluation is not new. As you already know, the earliest forms of evaluation took place, centuries ago, in many cultures including ancient Egypt.

A more modern discipline of evaluation emerged from social science research. Evaluation has been frequently used to assess educational, health and social programs. More recently, evaluation has been used to look at industry, housing, criminal justice, mental health, work productivity, social security, agricultural programs, job and occupational training, and community development. Development evaluation has emerged as a sub-discipline of evaluation.

The notion of development has changed over the past decades:

- 1950's: after World War II, were characterized by a focus on rebuilding, reconstruction, technical assistance, and engineering.

- 1960's: as many newly independent countries were created, the development world placed primary emphasis on economic growth, financing, and the creation of projects in the hopes that higher economic growth would lift more people out of poverty.

- 1970's: the emphasis shifted again to the social sectors or basic needs—education, health, and social welfare. As such, the development community began to do longer term planning and to make social sector investments.

- 1980's: further shifts toward structural adjustment policies and lending. Adjustment lending was used to support major policy reforms and to help countries cope with financial and debt crises. Adjustment lending was linked to specific conditionality.

- 1990's: the focus shifted to country-level assistance, that is, more comprehensive country-based programs as opposed to an emphasis on individual projects. More emphasis was given to building capacity and institutions within developing countries.

As we move into the first decade of the 21st century, trends in development highlight poverty reduction, partnerships, participation, and a results orientation. Sector-wide, country level, and global levels are now being increasingly used.

> Sector-wide approaches, or SWAPs, support a country-led program for a coherent sector in a comprehensive and coordinator manner. It is characterized by: (i) country-owned sector policies and strategies; (ii) sustained, country-led partnership among development partners and key stakeholders; (iii) common programs and an expenditure framework based on agreed priorities and strategies; (iv) common approaches in planning, implementation and *monitoring and evaluation*; and (v) increasing reliance on country's systems and procedures. [38]

Thus, SWAPs also include new challenges for funding strategies and donor partnership/coordination at the broader sectoral level.

The current decade has also seen a shift from project to country level assistance programs and national poverty reduction strategies; that is, from partial to more comprehensive development approaches.

Table 1.3 summarizes these transitions.

Table 1.3: Changing Development Concepts

Decades	Objectives	Approaches	Discipline
1950's	Reconstruction	Technical assistance	Engineering
1960s	Growth	Projects	Finance
1970s	Basic needs	Sector investment	Planning
1980s	Adjustment	Adjustment lending	Neoclassical economics
1990s	Capacity-building	Country assistance	Multi-disciplinary
2000s	Poverty reduction	Partnerships	Results-based management

Source: Robert Picciotto, PowerPoint Presentation. World Bank 2002

---

[38] "Sector-wide Approach (SWAp),"
http://www.worldbank.org/opcil/SWAps.html

The Evolution of Development Evaluation

Development evaluation is no longer an approach that views the program from the detached perspective of an outsider, but now rather as a participatory, collaborative endeavor. The role of the evaluator has also expanded more: from evaluator as accountant to evaluator as researcher to evaluator (especially the internal evaluator) as facilitator (see Figure 1.1). Evaluators are now expected to have a broader and more diverse skill set.[39] The relationships between participants, donors, and evaluators are also changing. Where once evaluations were top-down events, they are changing to a more collaborative, joint approach that brings the key stakeholders together in designing and carrying out the evaluation.
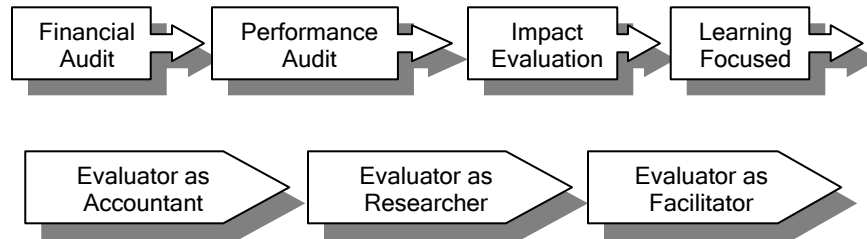

Fig. 1.1: Evolution of Development Evaluation, Expanding Roles

Within the development context, the official OECD definition of development evaluation is as follows:

> The systematic and objective assessment of an on-going or completed project, program or policy, its design, implementation and results. The aim is to determine the relevance and fulfillment of objectives, development efficiency, effectiveness, impact, and sustainability. An evaluation should provide information that is credible and useful, enabling the incorporation of lessons learned into the decision-making process of both recipients and donors.

Evaluation also refers to the process of determining the worth or significance of an activity, policy or program; an assessment, as systematic and objective as possible, of a planned, on-going, or completed intervention. (See also *OECD Glossary of Key Terms in Evaluation and Results-Based Management* in your supplementary materials.)

---

[39] R.C. Sonnichsen, *High impact internal evaluation* (Thousand Oaks, CA: Sage, 2000).

A wide variety of methodologies and practices have been utilized in the development evaluation community. It has generally become accepted that a mix of theories, analysis strategies, and methodologies often works best in development evaluation, especially given the growing scale and complexity of development projects/programs/policies. Mixing approaches can help strengthen the evaluation. This mix of methods is called **triangulation**, which refers to:

> the use of several theories, sources or types of information, and/or types of analysis to verify and substantiate an assessment. By combining multiple data sources, methods, analyses, or theories, evaluators seek to overcome the bias that comes from single informants, single methods, single observer, or single theory studies.[40]

Development evaluation is built on OECD criteria and relies heavily on triangulation. Triangulation can help to synergize complementarity among multiple methods and practices.

> We have learned that every social science method has important weaknesses, but also that the existence of such weaknesses is not fatal. What they imply is a need to use several methods together, so that the strengths of one can compensate for the limitations of the other. We think less today about the absolute merits of one method versus another and more about whether and how using them in concert could results in more conclusive findings.[41]

This trend is likely to persist for some time to come.

> Evaluation continues to become ever more methodologically diverse. It is by now well established that the full array of social science methods belongs in the evaluator's methodological toolkit—tools from psychology, statistics, education, sociology, political science, anthropology, and economics.[42]

---

[40] OECD, *Glossary of key terms in evaluation and results based management.* Paris: OECD. p 37.

[41] Eleanor Chelimsky (1995). "Preamble: New dimensions in evaluation," in *Evaluating country development policies and programs: New approaches for a new agenda,* Number 67, Fall 1995. Robert Picciotto, Ray C. Rist, editors. San Francisco: Jossey-Bass Publishers (Publication of the American Evaluation Association). p. 7.

[42] E. Chelimsky and W.R.Shadish (1997). *Evaluation for the 21st century: A handbook.* Thousand Oaks, CA: Sage Publications.

Ultimately, the choice of which evaluation design and methodology – or combination of designs and methodologies – to use will be determined by the questions being asked and the information sought.

Built on OECD DAC Criteria

Development evaluation is built on the Organisation for Economic Co-operation and Development (IOECD) Development Co-operation Directorate (DAC) criteria. In 1991, broad development evaluation principles were endorsed and updated by the Organization for Economic Cooperation and Development (OECD) Development Assistance Committee (DAC). The DAC Network on Development Evaluation brings together 30 bilateral and multilateral development agencies. They work on improving evaluation for more effective development assistance.

For our purposes, we shall focus on the five DAC criteria for evaluating development assistance.

DAC Criteria for Evaluating Development Assistance

- **Relevance**: The extent to which the aid activity is suited to the priorities and policies of the target group, recipient, and donor.

- **Effectiveness**: A measure of the extent to which an aid activity attains its objectives.

- **Efficiency**: Efficiency measures the outputs - qualitative and quantitative - in relation to the inputs. It is an economic term, which signifies that the aid uses the least costly resources possible in order to achieve the desired results. This generally requires comparing alternative approaches to achieving the same outputs, to see whether the most efficient process has been adopted.

- **Impact**: The positive and negative changes produced by a development intervention, directly or indirectly, intended or unintended. This involves the main impacts and effects resulting from the activity on the local social, economic, environmental and other development indicators. The examination should be concerned with both intended and unintended results and must include the positive and negative impact of external factors, such as changes in terms of trade and financial conditions.

- **Sustainability**: Sustainability is concerned with measuring whether the benefits of an activity are likely to continue after donor funding has been withdrawn. Projects need to be environmentally as well as financially sustainable.[43]

---

[43] OECD, DAC Criteria for Evaluating Development Assistance, http://www.oecd.org/document/22/0,2340,en_2649_34435_2086550_1_1_1_1,00.html

See the Annex for full text of DAC Criteria for Evaluating Development Assistance.

Principles and Standards

The DAC developed specific principles for the evaluation of development assistance that address the following issues:

- the purpose of evaluation
- impartiality and independence
- credibility
- usefulness
- participation of donors and recipients
- donor cooperation
- evaluation programming
- design and implementation of evaluations
- reporting, dissemination and feedback
- application of these principles.[44]

A review of the DAC Principles for Evaluation of Development Assistance[45] was conducted in 1998. The review is comparing the DAC standards to those of other organizations and looking for consistency and possible areas to expand for the DAC standards. Members responded with recommendations for possible revisions to the Principles, including:

- modifying the statement of purpose …
- directly addressing the question of decentralized evaluations and participatory evaluations
- elaborating more on the Principles and practices for recipient participation and donor cooperation
- introducing recent developments in evaluation activity such as performance measure[ment], [status], and success rating systems, and developing a typology of evaluation activity.[46]

---

[44] See OECD DAC, Principles for Evaluation of Development Assistance, http://www.oecd.org/dataoecd/31/12/2755284.pdf

[45] See Annexes.

[46] "Review of the DAC Principles for Evaluation of Development Assistance," 1998, p. 18, http://www.oecd.org/dataoecd/31/12/2755284.pdf

Why do we need principles and standards? They:

- promote accountability
- facilitate comparability
- enhance reliability and quality of services provided.[47]

In 1994 the Program Evaluation Standards were published in the United States and were approved by the American National Standards Institute (ANSI) as the American National Standards for program evaluation (see Further Reading and Resources, later in this module). This document defines evaluation as the systematic investigation of the worth or merit of an object (a program, project, or product). For example, a development program may be very effective in improving a country's economy (merit or performance), but only at the expense of the welfare of its people (intrinsic worth). Both merit and worth are important in evaluating development programs.

The Program Evaluation Standards were updated in 1998. These updated standards can be found at:

http://www.eval.org/EvaluationDocuments/progeval.html

The United Nations has also established Norms for Evaluation in the UN System. They can be found at the following website:

http://www.uneval.org/docs/ACFFC9F.pdf

The UN has also developed Standards for Evaluation in the UN System. The standards can be found at the following website:

http://www.uneval.org/docs/ACFFCA1.pdf

The American Evaluation Association (AEA) has also established Guiding Principles. They were updated in July of 2004. They can be found at the following website:

http://www.eval.org/Publications/GuidingPrinciples.asp

---

[47] Robert Picciotto. (2005) "The value of evaluation standards: A comparative assessment." in *Journal of Multidisciplinary Evaluation.* Number 3. Available online at
http://evaluation.wmich.edu/jmde/content/JMDE003content/PDFs%20JMDE%20003/4_%20The_Value_of_Evaluation_Standards_A_Comparative_Assessment.pdf

In March of 2006, the DAC Evaluation Network established the *DAC Evaluation Quality Standards[48]*. They are currently used on a trial basis, for test phase application.

These standards identify the key pillars needed for a quality evaluation process and product. They are intended to:

- provide standards for the process (conduct) and products (outputs) of evaluations
- facilitate the comparison of evaluations across countries (meta-evaluation)
- facilitate partnerships and collaboration on joint evaluations
- better enable member countries to make use of each others' evaluation findings and reports (including good practice and lessons learned)
- streamline evaluation efforts.

The Ten Parts of the Evaluation Quality Standards:

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

---

[48] DAC Evaluation Quality Standards (for test phase application). 30-31 March 2006. at: http://www.oecd.org/dataoecd/30/62/36596604.pdf also available in the Annexes.

These new DAC Evaluation Quality Standards can be found at the following website:

http://www.oecd.org/dataoecd/30/62/36596604.pdf

Principles and standards are a big part of ethical considerations and will be discussed further in Module 15, *Evaluation Ethics, Standards, and Guiding Principles.*

Evaluation and Independence

A critical component of credibility of evaluation is **independence**. The OECD Glossary defines and independent evaluation this way. "An evaluation carried out by entities and persons free of the control of those responsible for the design and implementation of the evaluation." The definition also includes the following note. "The credibility of an evaluation depends in part on how independently it has been carried out. Independence implies freedom from political influence and organizational pressure. It is characterized by full access to information and by full autonomy in carrying out investigations and reporting findings". [49]

However, independence is not isolation. When in-house evaluators are used, the interaction between evaluators, program managers, staff, and beneficiaries enhances the evaluation. The evaluation can be conducted internally, as long as the evaluators are not involved in making decisions about the design and implementation of the evaluation.

The Evaluation Cooperation Group (ECG) has identified four dimensions or criteria of evaluation independence. They are:

- organizational independence
- behavioral independence
- avoidance of conflicts of interest
- protection from external influence.

The heads of evaluation of the Multilateral Development Banks (MDB) use the information in Table 1.2 to review their level of independence:

---

[49] OECD (2002). *Glossary of key terms in evaluation and results based management.* Paris: OECD Publications. p. 25.

Table 1.2: Rating Template for Governance of Evaluation Organizations. Evaluation Cooperation Group of the Multilateral Development Banks, 2002

Criterion	Aspects	Indicators
I. Organizational independence	The structure and role of the evaluation unit	Whether the evaluation unit has a mandate statement that makes clear that its scope of responsibility extends to all operations of the organization, and that its reporting line, staff, budget and functions are organizationally independent from the organization's operational, policy, and strategy departments and related decision-making
	The unit is accountable to, and reports evaluation results to, the head or deputy head of the organization or its governing Board	Whether there is a direct reporting relationship between the unit and the Management and/or Board of the institution
	The unit is located organizationally outside the staff or line management function of the program, activity or entity being evaluated	The unit's position in the organization relative to the program, activity or entity being evaluated
	The unit reports regularly to the larger organization's audit committee or other oversight body	Reporting relationship and frequency of reporting to the oversight body
	The unit is sufficiently removed from political pressures to be able to report findings without fear of repercussions	Extent to which the evaluation unit and its staff are not accountable to political authorities, and are insulated from participation in political activities
	Unit staffers are protected by a personnel system in which compensation, training, tenure and advancement are based on merit	Extent to which a merit system covering compensation, training, tenure and advancement is in place and enforced
	Unit has access to all needed information and information sources	Extent to which the evaluation unit has unrestricted access to the organization's staff, records, co-financiers and other partners, clients, and those of programs, activities or entities it funds or sponsors

Criterion	Aspects	Indicators
II. Behavioral Independence	Ability and willingness to issue strong, uncompromising reports	Extent to which the evaluation unit has issued reports that invite public scrutiny (within appropriate safeguards to protect confidential or proprietary information and to mitigate institutional risk) of the lessons from the organization's programs and activities; propose standards for performance that are in advance of those in current use by the organization; and critique the outcomes of the organization's programs, activities and entities
	Ability to report candidly	Extent to which the organization's mandate provides that the evaluation unit transmits its reports to the Management/Board after review and comment by relevant corporate units but without management-imposed restrictions on their scope and comments
	Transparency in the reporting of evaluation findings	Extent to which the organization's disclosure rules permit the evaluation unit to report significant findings to concerned stakeholders, both internal and external (within appropriate safeguards to protect confidential or proprietary information and to mitigate institutional risk).
III. Protection from outside interference	Proper design and execution of an evaluation	Extent to which the evaluation unit is able to determine the design, scope, timing and conduct of evaluations without Management interference
	Evaluation study funding	Extent to which the evaluation unit is unimpeded by restrictions on funds or other resources that would adversely affect its ability to carry out its responsibilities
	Judgments made by the evaluators	Extent to which the evaluator's judgment as to the appropriate content of a report is not subject to overruling or influence by an external authority
	Evaluation unit head hiring/firing, term of office, performance review and compensation	Mandate or equivalent document specifies procedures for the hiring, firing, term of office, performance review and compensation of the evaluation unit head that ensure independence from operational management
	Staff hiring, promotion or firing	Extent to which the evaluation unit has control over staff hiring, promotion, pay increases, and firing, within a merit system
	Continued staff employment	Extent to which the evaluator's continued employment is based only on reasons related to job performance, competency or the need for evaluator services

Criterion	Aspects	Indicators
IV. Avoidance of conflicts of interest	Official, professional, personal or financial relationships that might cause an evaluator to limit the extent of an inquiry, limit disclosure, or weaken or slant findings	Extent to which there are policies and procedures in place to identify evaluator relationships that might interfere with the independence of the evaluation; these policies and procedures are communicated to staff through training and other means; and they are enforced
	Preconceived ideas, prejudices or social/political biases that could affect evaluation findings	Extent to which policies and procedures are in place and enforced that require evaluators to assess and report personal prejudices or biases that could imperil their ability to bring objectivity to the evaluation; and to which stakeholders are consulted as part of the evaluation process to ensure against evaluator bias
	Current or previous involvement with a program, activity or entity being evaluated at a decision-making level, or in a financial management or accounting role; or seeking employment with such a program, activity or entity while conducting the evaluation	Extent to which rules or staffing procedures that prevent staff from evaluating programs, activities or entities for which they have or had decision-making or financial management roles, or with which they are seeking employment, are present and enforced
	Financial interest in the program, activity or entity being evaluated	Extent to which rules or staffing procedures are in place and enforced to prevent staff from evaluating programs, activities or entities in which they have a financial interest
	Immediate or close family member is involved in or is in a position to exert direct and significant influence over the program, activity or entity being evaluated	Extent to which rules or staffing procedures are in place and enforced to prevent staff from evaluating programs, activities or entities in which family members have influence

Sources: U.S. General Accounting Office, *Government Auditing Standards*, Amendment 3 (2002); OECD/DAC, *Principles for Evaluation of Development Assistance* (1991); INTOSAI, *Code of Ethics and Auditing Standards* (2001); Institute of Internal Auditors, *Professional Practices Framework* (2000); European Federation of Accountants, *The Conceptual Approach to Protecting Auditor Independence* (2001); Danish Ministry of Foreign Affairs, *Evaluation Guidelines* (1999); Canadian International Development Agency, *CIDA Evaluation Guide* (2000).

Growth of Professional Evaluation Associations

Jean Serge Quesnel[50] has written on the importance of evaluation associations and networks. He identifies the growth of new organizations and associations. He also points out that the growth has become more accelerated in the past few years.

The growth in the number of emerging evaluation associations, groups, and networks shows that evaluation is becoming increasingly valued.

Professional evaluation associations create a support system and allow for professionalism within the evaluation community. This support contributes to capacity development in development evaluation.

Another notable development is the number of new national, regional, and international evaluation associations that have sprung up around the world. As of 2005, there are *more than fifty* such evaluation associations in developing and developed countries alike.[51]

At the national level, to name just a few, the Canadian Evaluation Society was established in 1981, the Uganda Evaluation Society in 2002, and the Malaysia Evaluation Society in 1999.

On the regional level, the European Evaluation Society was established in 1994, the Australasian Evaluation Society in 1991, and the African Evaluation Association in 1991.

---

[50] J.S. Quesnel (2006). "The Importance of Evaluation Associations and Networks" UNICEP Regional Office for CEE/CIS and IPEN Issue #5 *New Trends in Development Evaluation.* Article pp 17-25:: at : http://www.unicef.org/ceecis/New_trends_Dev_EValuation.pdf

[51] See Annex I.5, and "Digital Resources for Evaluators," http://www.resources4evaluators.info/CommunitiesOfEvaluators.html

An important international organization for evaluation is the International Organisation for Cooperation in Evaluation (IOCE). IOCE. It is a loose alliance of regional and national evaluation organisations (associations, societies and networks) from around the world. Their aim is collaborating to:

- build evaluation leadership and capacity in developing countries

- foster the cross-fertilisation of evaluation theory and practice around the world.

assist the evaluation profession to take a more global approach to contributing to the identification and solution of world problems.[52]

Another important international organization for evaluation is the International Development Evaluation Association (IDEAS)[53]. IDEAS, was created in 2001 to help build evaluation capacity in developing countries. IDEAS mission is "to advance and extend the practice of development evaluation by refining methods, strengthening capacity, and expanding ownership". IDEAS strategy is to:

- Promote development evaluation for results, transparency and accountability in public policy and expenditure

- Give priority to evaluation capacity development (ECD)

- Foster the highest intellectual and professional standards in development evaluation…"

---

[52] International Organisation for Cooperation in Evaluation (IOCE) http://ioce.net/overview/general.shtml

[53] International Development Evaluation Association (IDEAS), http://www.ideas-int.org/Index.aspx

IDEAS aims to be:

- global – serving development evaluators everywhere;

- voluntary – membership, participation and professional contributions are voluntary

- inclusive – membership open to all who are interested in development evaluation

- democratic – elected leadership

- educational – advancing evaluation for the benefit of all through education and training

- pluralistic – encouraging multiple theories, methods, and perspectives

- responsive – assessing and meeting the needs of development evaluators

- participatory – all members encouraged to participate actively.

IDEAS is striving to attract committed members world-wide (but particularly from developing countries and transition economies), who will:

- promote development evaluation for results, transparency and accountability in public policy and expenditure

- give priority to evaluation capacity development (ECD)

- foster the highest intellectual and professional standards in development evaluation

- refrain from activities that other organisations are better placed to undertake

- nurture partnerships with like-minded organisations

- encourage national and regional development evaluation groups.

Summary

In this module you were introduced to development evaluation. Use the following checklist of information that you should know to help you review this module.

- ❑ define evaluation

- ❑ define development evaluation

- ❑ describe the purposes and uses of evaluation

- ❑ discuss the history of evaluation and development evaluation

- ❑ discuss the origins of evaluation

- ❑ describe the changes in development evaluation since the 1950's

- ❑ describe evaluation independence

- ❑ discuss OECD DAC criteria for evaluating development assistance

  – relevance

  – effectiveness

  – efficiency

  – impact

  – sustainability

Quiz Yourself

Answer the following multiple-choice questions to help test your knowledge of the material in Module #1.

You will find the answers to the questions on the last page of this module.

1. Which of the following is the **OECD definition of evaluation**?
   a. the action of appraising or valuing (goods, etc); a calculation or statement of value
   b. a standard by which a person or thing may be gauged
   c. the process of determining the worth or significance of an activity, policy or program
   d. to determine the value of a thing

2. Which of the following are the **four distinct purposes of evaluation** as given in this module?
   a. ethical, sustainability, decisional, and motivational
   b. cost effectiveness, sustainability, anti-corruption, and ethical
   c. ethical, managerial, decisional, and educational/motivational

3. Which of the following is the **definition of formative evaluation** given in this module?
   a. evaluations intended to improve performance, and are conducted during the implementation phase of projects or programs
   b. studies conducted at the end of an intervention or a phase of that intervention to determine the extent to which anticipated outcomes were produced
   c. synthesis monitoring and evaluating information from earlier studies to assess the likely outcomes of proposed new projects

4. Which of the following is the **definition of summative evaluation** given in this module?
   a. synthesis monitoring and evaluating information from earlier studies to assess the likely outcomes of proposed new projects
   b. evaluations intended to improve performance, and are most often conducted during the implementation phase of projects or programs
   c. studies conducted at the end of an intervention or a phase of that intervention to determine the extent to which anticipated outcomes were produced

5. **Match the decade** in the left column **with the Objectives, Approaches, and Disciplines** in the right column by drawing a line between the ones that match.

Decade	*Objective*	*Approach*	*Discipline*
1950's	Capacity-building	Country assistance	Multi-disciplinary
1960s	Adjustment	Adjustment lending	Neoclassical economics
1970s	Reconstruction	Technical assistance	Engineering
1980s	Basic needs	Sector investment	Planning
1990s	Poverty reduction	Partnerships	Results-based management
2000s	Growth	Projects	Finance

6. List the kinds **activities that evaluators generally perform**.

7. Which of the following lists the **DAC criteria for evaluating development assistance**?
   a. relevance, effectiveness, efficiency, impact, and sustainability
   b. efficiency, effectiveness, utility, accuracy, and impact
   c. utility, effectiveness, efficiency, and competence
   d. relevance, effectiveness, propriety, impact, and sustainability

Reflection

Consider what you have learned about the history and development of evaluation.

- What skills do you have that will help you in development evaluation?

- What skills do you think you will need to improve to become a better evaluator?

- How has the history of evaluation affected development?

- What new changes do you see that may affect evaluation in the future?

- What changes do you predict for evaluation in the future?

- What is the importance of principles and standards to a development evaluation?

Application Exercise 1.1

1. Imagine that you have been asked to justify why development evaluation should be a budgeted expense for a new national program. The program was designed to improve the education of families about effective health practices. What would you say in defense of development education?

2. Interview an evaluator in your field (or review recent evaluation reports conducted in your field) to determine the extent to which standards and guiding principles are addressed in evaluations that this individual has seen. Where do the strengths seem to be? The weaknesses? Share your findings with evaluation colleagues and listen to their comments and experiences. Do you see any patterns?

Further Reading and Resources

Fitzpatrick, Jody. L.; James R. Sanders, and Blaine R. Worthen (2004). *Program Evaluation: Alternative Approaches and Practical Guidelines.* New York: Pearson Education Inc.

Feuerstein, M. T. (1986). *Partners in Evaluation: Evaluating Development and Community Programs with Participants.* London: MacMillan, in association with Teaching Aids At Low Cost.

Furubo, Jan-Eric and Sandahl, R. (2002). "Coordinated Pluralism," in *International Atlas of Evaluation*, Jan-Eric Furubo, Ray Rist, and Rolf Sandahl, editors, New Brunswick, NJ: Transaction Publishers.

Furubo, Jan-Eric,: Ray Rist and Rolf Sandahl, editors (2002). *International Atlas of Evaluation.* New Brunswick, NJ: Transaction Publishers.

Lawrence, J. (1989). Engaging Recipients in Development Evaluation—the "Stakeholder" Approach. *Evaluation Review, 13*(3).

Rossi, Peter and Howard Freeman (1993). *Evaluation: A Systematic Approach.* Thousand Oaks, CA: Sage Publications.

Patton, M.Q. (1997). *Utilization-focused Evaluation* (3rd ed.). Thousand Oaks, CA: Sage Publications.

Valdez, Joseph and Michael Bamberger (1994). *Monitoring and Evaluating Social Programs in Developing Countries: A Handbook for Policymakers, Managers and Researchers.* World Bank Economic Development Institute Development Studies.

Molund, Stefan and Göran Schill (2004). *Looking Back, Moving Forward: SIDA Evaluation Manual.* Stockholm SIDA.
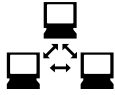
Websites

Evaluation Organizations

Australasian Evaluation Society: http://www.aes.asn.au/

African Evaluation Association: http://www.geocities.com/afreval/

American Evaluation Association: www.eval.org

Canadian Evaluation Association: www.evaluationcanada.ca

European Evaluation Society: www.europeanevaluation.org

German Society for Evaluation (DeGEval) standards: http://www.degeval.de/

International Development Evaluation Association (IDEAS) http://www.ideas-int.org/

IDEAS website with links to many organizations http://www.ideas-int.org/Links.aspx

International Organisation for Cooperation in Evaluation http://ioce.net/overview/general.shtml

Italian Evaluation Association guidelines: http://www.valutazioneitaliana.it/

Swiss Evaluation Society (SEVAL) standards: http://seval.ch/.

Evaluation Standards

American Evaluation Association Guiding Principles http://www.eval.org/Publications/GuidingPrinciples.asp

African Evaluation Association Evaluation Standards and Guidelines: http://www.afrea.org/

Australasian Evaluation Society ethical guidelines for evaluators: http://www.aes.asn.au/content/ethics_guidelines.pdf.

DAC Evaluation Quality Standards (for test phase application). 30-31 March 2006.: http://www.oecd.org/dataoecd/30/62/36596604.pdf

German Society for Evaluation (DeGEval) standards: http://www.degeval.de/standards/standards.htm

Italian Evaluation Association guidelines: http://www.valutazioneitaliana.it/statuto.htm#Linee

Program Evaluation Standards updated in 1998 http://www.eval.org/EvaluationDocuments/progeval.html

Swiss Evaluation Society (SEVAL) standards: http://seval.ch/

United Nations Evaluation Group (UNEG) Norms for Evaluation in the UN System.
http://www.uneval.org/docs/ACFFC9F.pdf

M&E In's and Out's. *InsideOut,* Issue #3: October/November 2005.
http://www.insideoutresearch.co.za/docs/newsletter3.pdf

United Nations Evaluation Group (UNEG) Standards for Evaluation in the UN System.
http://www.uneval.org/docs/ACFFCA1.pdf

Evaluation Methods References

DFID on SWAps http://www.keysheets.org/red_7_swaps_rev.pdf

Evaluation Center, Western Michigan University:
http://www.wmich.edu/evalctr/

Human Rights Education:
www.hrea.org/pubs/EvaluationGuide/

Linkages Between Audit and Evaluation in Canadian Federal Developments," Treasury Board of Canada:
http://www.tbs-sct.gc.ca/pubs_pol/dcgpubs/TB_h4/evaluation03_e.asp

Monitoring and Evaluation of Population and Health Programs, MEASURE Evaluation Project, University of North Carolina at Chapel Hill:
http://www.cpc.unc.edu/measure

National Aeronautics and Space Act of 1958:
http://www.hq.nasa.gov/office/pao/History/spaceact.html

OECD DAC, Principles for Evaluation of Development Assistance, http://www.oecd.org/dataoecd/31/12/2755284.pdf

OECD, DAC Criteria for Evaluating Development Assistance,
http://www.oecd.org/document/22/0,2340,en_2649_34435_2086550_1_1_1_1,00.html

Participatory Monitoring and Evaluation: Learning from Change, IDS Policy Briefing, issue 12, November 1998,
http://www.ids.ac.uk/ids/bookshop/briefs/brief12.html

Picciotto, Robert. (2005) "The value of evaluation standards: A comparative assessment." in *Journal of Multidisciplinary Evaluation.* Number 3. Available online at
http://evaluation.wmich.edu/jmde/content/JMDE003content/PDFs%20JMDE%20003/4_%20The_Value_of_Evaluation_Standards_A_Comparative_Assessment.pdf

Proposal for Sector-wide Approaches (SWAp)
http://enet.iadb.org/idbdocswebservices/idbdocsInternet/IADBPublicDoc.aspx?docnum=509733

The Institute of Internal Auditors,
http://www.theiia.org

The International Organization of Supreme Audit Institutions
http://www.gao.gov/cghome/parwi/img4.html

UNFPA List of Evaluation Reports and Findings. United Nations
Population Fund. Online:
http://www.unfpa.org/publications/index.cfm

United Nations Development Project Evaluation Office:
www.undp.org/eo/

UNICEP Regional Office for CEE/CIS and IPEN Issue #5
(2006). *New Trends in Development Evaluation.*:
http://www.unicef.org/ceecis/New_trends_Dev_EValuation.pdf

World Bank: www.worldbank.org

The World Bank Participation Sourcebook. (HTML format):
http://www.worldbank.org/wbi/sourcebook/sbhome.htm

Weiss, Carol (2004). Identifying the intended use(s) of an
evaluation. The IDRC.

http://www.idrc.ca/ev_en.php?ID=58213_201&ID2=DO_TOPIC p 1

Answers to Quiz Yourself

1. c

2. c

3. a

4. c

5.

1950's	Capacity-building	Country assistance	Multi-disciplinary
1960's	Adjustment	Adjustment lending	Neoclassical economics
1970's	Reconstruction	Technical assistance	Engineering
1980's	Basic needs	Sector investment	Planning
1990's	Poverty reduction	Partnerships	Results-based management
2000's	Growth	Projects	Finance

6.

- consult with all stakeholder

- manage evaluation budgets

- perform or conduct the evaluation or hire and contract staff to perform the evaluation

- identify standards for effectiveness (relying on authorizing documents or other sources)

- collect, analyze, interpret, and report on data and findings.

7. a

To continue on to this
Module's Presentation Slides click here

To continue on to the
Next Module click here

To return to the
Table of Contents click here