



Treasury Board of Canada

Conseil du Trésor du Canada

Principles for
the Evaluation
of Programs by
Federal Departments
and Agencies

Canada

**Principles for
the Evaluation
of Programs by
Federal Departments
and Agencies**

© Minister of Supply and Services Canada 1983

Cat. No. BT66-2/1981

ISBN 0-662-51594-3

Reprinted December 1984

TABLE OF CONTENTS

	Page No.
PREFACE	1
- Background	1
- Purpose	1
- Scope	2
- Definitions	2
CHAPTER 1 CONDUCTING PROGRAM EVALUATIONS	3
1.1 Overview	3
1.2 Evaluation Assessment (Pre-Evaluation Planning)	3
1.3 Evaluation	5
1.4 Decision-Making	5
CHAPTER 2 THE EVALUATION ASSESSMENT STUDY	7
2.1 Principles for the Conduct of an Evaluation Assessment Study	7
2.1.1 Understanding the Program and its Environment	8
2.1.2 Identification of the Users and Purposes of the Evaluation Study	11
2.1.3 Identification of Potential Evaluation Questions	12
2.1.4 Determination of Evaluation Approaches	13
2.1.5 Estimation of Costs and Resources	18
2.1.6 Determination of Evaluation Options	19
2.1.7 Formulation of Recommendations	20
2.2 Quality Characteristics of an Evaluation Assessment Study	20
CHAPTER 3 THE EVALUATION STUDY	25
3.1 Principles for the Conduct of an Evaluation Study	25
3.1.1 Collection of Information	26
3.1.2 Preparation of the Information	28
3.1.3 Analysis of the Information	29

	Page No.
3.1.4 Formulation Of Conclusions	33
3.1.5 Formulation of Recommendations	33
3.2 Quality Characteristics of an Evaluation Study	33
 CHAPTER 4 TERMS OF REFERENCE	 37
4.1 Contents of Terms of Reference	37
4.2 The Detailed Work Plan	38
4.2.1 Work Plan for the Evaluation Assess- ment study	38
4.2.2 Work Plan for the Evaluation Study	38
4.3 Authorization of Terms of Reference	40
4.4 Distribution of Terms of Reference	40
 CHAPTER 5 EVALUATION REPORTING	 41
5.1 Form	41
5.2 Content	41
5.3 Debriefing	43
5.4 Distribution of Reports	44

FOREWORD

This is a companion document to the newly published *Guide to the Program Evaluation Function*, which established guidelines for the implementation of program evaluation in the federal public service.

The Principles for the Evaluation of Programs deals with the actual process of a program evaluation. It offers principles for conducting and reporting on evaluations, intended to assist departmental program evaluation officers as they address the concerns expressed by deputy ministers and Cabinet Committees about a program, its performance, results and cost-effectiveness. Also included are quality characteristics, developed in cooperation with departmental program evaluation officers, for evaluation assessment studies and for evaluation studies.

The principles set down here should not be looked upon as substitutes for common sense, or for the ability to select the correct approach to a program evaluation. They reflect, indeed, the view that evaluation, like any other management function, is an exercise in professional judgment.

H.G. Rogers,
Comptroller General of Canada.

PREFACE

BACKGROUND

The Treasury Board Policy 1977-47, "Evaluation of Programs by Departments and Agencies", recognizes that program evaluation is an integral part of the managerial responsibilities of the deputy heads of departments and agencies. A program evaluation questions the rationale of a program, and provides information on its performance, results and cost-effectiveness. The findings and recommendations of such evaluations can be used by deputy heads to:

- make more informed decisions on the management and resourcing of their programs;
- be accountable for the programs for which they are responsible; and
- provide quality advice to ministers.

As part of its responsibilities for effective program evaluation in the federal government, the Office of the Comptroller General has issued the *Guide on the Program Evaluation Function*, which covers the establishment and continuing operation of program evaluation in departments and agencies. The present document, *Principles for the Evaluation of Programs by Federal Departments and Agencies*, discusses the actual conduct of evaluations.

PURPOSE

The major purpose of this document is to provide departments and agencies with a useful framework for the planning, conduct and reporting of evaluations of programs. This document does not attempt to prescribe specific or detailed evaluation techniques; instead it identifies and discusses the factors that are useful to consider in an evaluation.

The second purpose of this document is to identify the important factors to be considered in assessing the quality of evaluations. These factors represent an initial agenda of the items that will be considered by the Office of the Comptroller General whenever it comments on the quality of selected evaluations. It may be more beneficial to develop explicit standards after the government has had more experience with the conduct and use of evaluation studies.

SCOPE

The principles outlined in this document concern the evaluations of programs carried out pursuant to the Treasury Board Policy Circular 1977-47 on the "Evaluation of Programs by Federal Departments and Agencies".

DEFINITIONS

The terms "program evaluation" and "evaluation" are used interchangeably in this document to refer to the evaluations of continuing programs carried out pursuant to Circular 1977-47.

The terms "program" and "program component"¹ are also used interchangeably, to refer to a group of activities - usually a subset of one Estimates program - which:

- has a common objective (or set of related objectives) established at the level of concern of deputy heads;
- contributes to the department's long-term objectives; and
- is of appropriate size and importance to be a focus of and support for program decision-making at the departmental level.

The term "program effectiveness" refers here to the degree to which a program:

- (i) makes sense in terms of the relationships between its activities and its expected results;
- (ii) achieves its objectives;
- (iii) produces intended and unintended results; and
- (iv) is cost-effective.

¹ For a detailed discussion of the concept of program component, see the Guide on the Program Evaluation Function, Office of the Comptroller General of Canada, Ottawa, 1981.

CHAPTER 1

CONDUCTING PROGRAM EVALUATIONS

This chapter recapitulates the main steps of the evaluation process discussed in the *Guide on the Program Evaluation Function*. Principles for the conduct and reporting of evaluations are presented in the following chapters.

1.1 OVERVIEW

The *Guide on the Program Evaluation Function* views the process of carrying out evaluations of programs as comprising three phases:

- pre-evaluation planning (evaluation assessment) to decide on the appropriate questions to be asked in the evaluation study and the approaches to be used for answering them;
- conducting and reporting on the evaluation study; and
- decision-making based on the evaluation findings and recommendations.

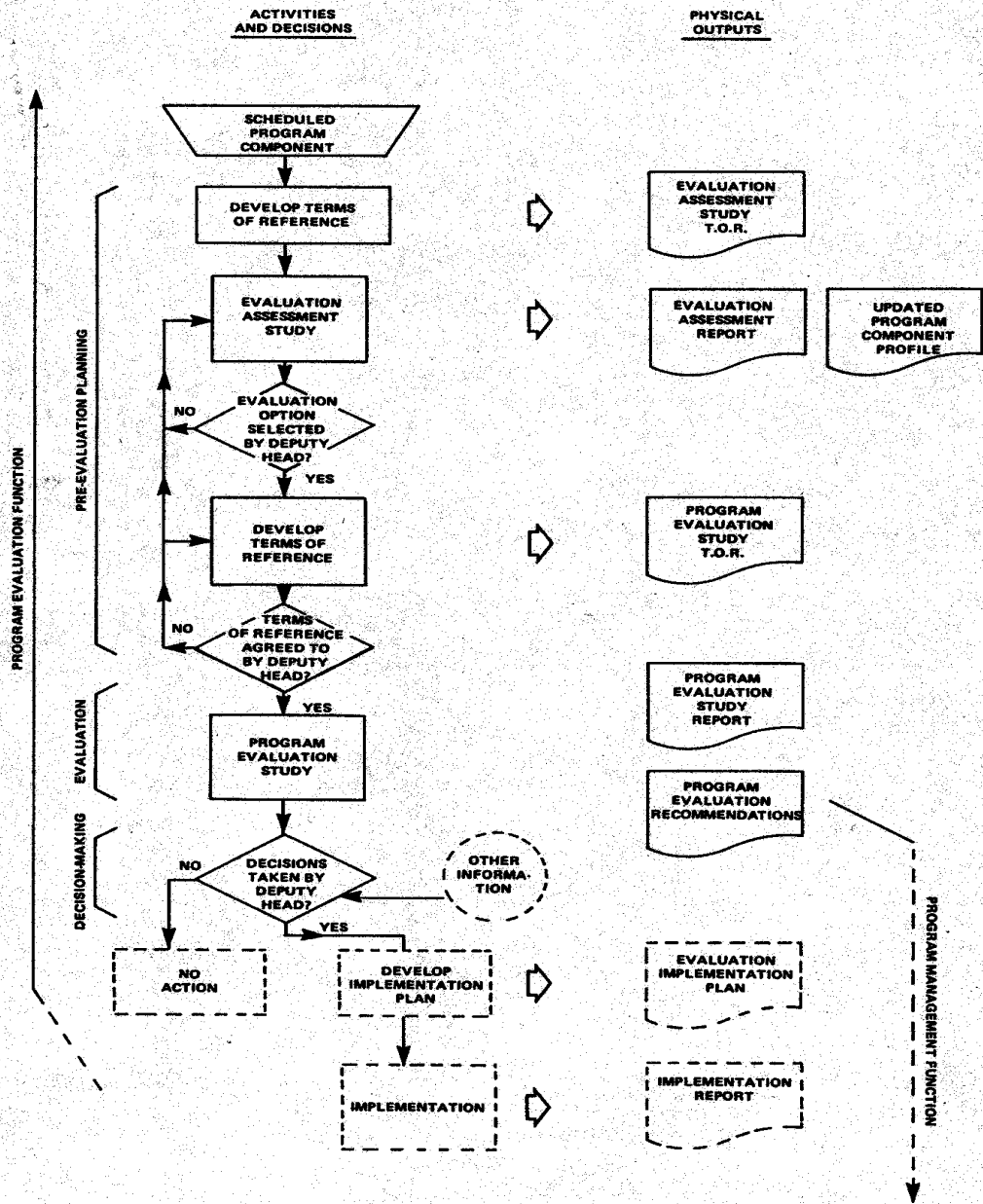
Figure 1.1 illustrates such a process, indicating both the activities and decisions that are normally taken as well as the documents typically produced. Each phase in this process will be briefly discussed in this chapter.

1.2 EVALUATION ASSESSMENT (Pre-Evaluation Planning)

Evaluation assessment is the front-end planning part of the evaluation of a program. It enables the client - the deputy head - to ensure an appropriate focus for the ensuing evaluation study. The degree of formality of the pre-evaluation planning phase will vary. In many cases, departments will find it useful to conduct an assessment study resulting in a report which will provide the deputy head with options as to the specific questions that could be answered in the evaluation study, along with related evaluation approaches and estimated costs. This is likely to be the case for large or complicated programs. In some cases, however, the assessment phase may be less formal, resulting in terms of reference for the evaluation study, but without an actual assessment study and report.

FIGURE 1.1

THE PROGRAM EVALUATION PROCESS: A REPRESENTATIVE MODEL



The conduct of an assessment study requires terms of reference as to what it will do and how it will do it, and the preparation of an evaluation assessment report. Chapter 2 of this document states the quality characteristics of an adequate assessment study and outlines the principles which are useful to follow in its conduct. Chapters 4 and 5 respectively describe the characteristics of good quality terms of reference and evaluation assessment reports.

1.3 EVALUATION

The evaluation study is the important operational phase during which the required information is collected, the analysis carried out, the findings produced, and the recommendations written. Principles for the conduct and reporting of evaluation studies are set out in Chapters 3 and 5 respectively. In addition, to help ensure that the questions asked in an evaluation study are adequately answered, terms of reference should be developed and approved by the client (Chapter 4).

1.4 DECISION-MAKING

This refers to the stage where decisions are taken as a result of the findings and recommendations of the evaluation study. As shown in figure 1.1, such decisions incorporate whatever other information is available on the program being considered, in addition to the findings and recommendations of the evaluation study. The implementation of program evaluation decisions typically involves the development of an implementation plan, which indicates what specific changes are to be made, how, and by whom, and an implementation report, which indicates what program changes occurred, at least in part, as a result of the decisions taken by the deputy head. As indicated in the *Guide on the Program Evaluation Function*, responsibility for following up actions taken as a result of any decisions reached by the deputy head should be clearly spelled out in the department's program evaluation policy.

Implementation of management decisions is viewed as a line management function; it is not further discussed in this document.

CHAPTER 2

THE EVALUATION ASSESSMENT STUDY

The purpose of an evaluation assessment study is to help ensure that the resources to be spent in an evaluation are used to answer the *right* questions in a credible manner. An evaluation assessment enables the deputy head to ensure the appropriate focus and approach for the ensuing evaluation study. As such, an evaluation assessment study should not ordinarily be lengthy or expensive, nor should it be a substitute for evaluation.

This chapter discusses the process and principles which can be usefully followed in the conduct of an assessment study. It also describes the characteristics of a good quality evaluation assessment study.

2.1 PRINCIPLES FOR THE CONDUCT OF AN EVALUATION ASSESSMENT STUDY

An evaluation assessment study should involve a number of tasks if it is to yield credible and relevant products. These usually are:

1. understanding the program and the environment in which it is operating;
2. identifying the users and purposes of the evaluation study;
3. identifying potential evaluation questions;
4. determining evaluation approaches;
5. estimating the time, costs and resources required for each evaluation approach;
6. determining specific options - questions and approaches -for the evaluation study;
7. formulating recommendations; and
8. reporting the findings and recommendations of the assessment study to the client and the senior managers responsible for the program under review.

Principles that are useful to take into account in the execution of each of these steps except the last, which is elaborated upon in Chapter 5, are discussed below.

2.1.1 Understanding the Program and its Environment

An important step in assessing the manner and extent to which a program may be evaluated is the development of a good understanding and documentation of the program, including the environment in which it is operating. Interaction between the evaluators and departmental program managers is essential in providing a good description of the program as it is actually operated. The process of understanding a program typically involves the following:

- defining the program;
- checking the consistency of the activities of the program with its mandate;
- determining the plausibility of the relationships between the activities of the program and the results it is intended to produce; and
- confirming the description of the program with the officials responsible for its management.

Defining the Program

A program is typically defined in terms of what it does, what it is supposed to do, how it is delivered, and the resources devoted to it. Table 2.1 outlines what information is required for a good understanding of a program for evaluation purposes. Such information can be collected from documentation and through interviews with informed personnel.

An evaluator first examines existing documents and records to develop an understanding of the purposes of the program, how it evolved throughout the years, how it is intended to operate, and the level and types of resources assigned to it. Examples of documents that are useful to review at the assessment phase are the legislative acts, Cabinet documents, policy papers, internal memoranda and operational reports, Treasury Board submissions, previous audit and evaluation reports, etc. The evaluator then conducts interviews to determine the perception of key people about the program. The choice of people to be interviewed depends on how comprehensive an understanding the evaluator wishes to develop about a

Table 2.1

PROGRAM COMPONENT PROFILE

Part A: Background

1. **Component Mandate:** A statement of the legal basis of the component and of what the program component must do and what it may do.
2. **Component Objective:** A statement of what impacts and effects the component is specifically designed to accomplish or contribute to.
3. **Component Description:** A short narrative explaining what the component entails: how it is delivered; the environment it operates in; the population served; and what it is to accomplish.
4. **Relation to Estimates Program:** The Estimates program or programs from which the component is funded should be identified and the relationship between the component's objective and that of the Estimates program explained.
5. **Component Resources:**
 - (i) **Fiscal Expenditures** - The operating, capital and grants and contribution costs of the component as well as the authorized person-years devoted to the component.
 - (ii) **Capital Assets** - An identification of the facilities and equipment in addition to the office space devoted to the component.

Part B: Elements and Structure

1. Component Elements

- (i) **Activities** - A list of the major work tasks and any powers or functions that characterize a given component and which are performed or administered by the component personnel.
- (ii) **Outputs** - A list of the goods and services which are produced or directly controlled by the component personnel and distributed outside the component organization, as well as any regulations or provisions in tax legislation produced by or monitored by component personnel.
- (iii) **Expected Impacts and Effects** - These are the further goods, services and regulations (if any) produced as a result of the component's outputs and the consequent expected chain of outcomes which occur outside the component on society or part thereof.

2. The Component's Structure: A description and chart showing the linkages between the component's elements, i.e., a program model.

program, the complexity of the program, and the time and resources available to carry out the assessment study. It is typical to interview senior departmental executives and program managers. It may also be useful to interview field staff if the program is largely decentralized, and the information collected at the centre is not sufficient to adequately describe the program.

Checking Consistency

In an evaluation assessment study, the activities of the program being evaluated should be checked for general consistency with its mandate. Significant divergences should be reported in writing to the deputy head. The deputy head may not want an evaluation to be conducted before the program is redirected to do what it is supposed to do, or may decide that an evaluation should still be conducted to determine the program's impacts and effects. If ascertaining the consistency of the activities of a program with its mandate appears to be a major endeavour, this could become the first topic to be investigated in the evaluation study.

Determination of Plausible Relationships

The relationships between the activities of a program and the results it is intended to produce must be assessed in terms of their plausibility, before the conduct of an evaluation study. In this way, it is possible to avoid the assessment of obviously unrealistic impacts and effects which would increase the cost of the evaluation without adding to the understanding of the program.

During an evaluation assessment study, the emphasis is on determining the *plausibility* rather than the *validity* of relationships between the program's activities and its intended results. The validity of these relationships is tested during the ensuing evaluation study. Hence, the analysis required at the evaluation assessment stage is based on common sense and on the best available knowledge in the appropriate field.

To determine the plausibility of these causal linkages, it is necessary to identify and assess issues such as:

- the appropriateness of the assumptions underlying the structure of the program;

- the adequacy of the type and magnitude of the program effort directed at the intended impacts and effects; and
- the likely significance of parallel programs or alternative explanations which may better account for the occurrence of the expected program results.

An assessment of the plausibility of the relationships between a program's activities and its expected results could become an evaluation topic, if a program is too complex or if little information is available on the program itself and on its relationship to other parallel programs.

Confirmation with Program Managers

The program structure and the analysis of the relationships between the activities of the program and its intended results should be reviewed with program managers to help ensure the relevance of the specific evaluation questions to be answered in the subsequent evaluation study, and hence increase its utility. Where divergences exist between documentation and program managers' views as to what a program should do, what it actually does, or how it is implemented, these varying views should be properly documented. The evaluation assessment study should take into account these varying views in the identification of potential evaluation issues and approaches, and in the determination of evaluation options.

The advantages and disadvantages of conducting an evaluation study based on several interpretations of the structure of a program should be explained to the client of the study. If more than one view is held of the program structure or its underlying assumptions, the conduct of an evaluation study based on only one interpretation of the program could produce irrelevant information. On the other hand, the investigation of the various views held of a program in an evaluation study may have serious cost implications. Therefore, it is important for the client of the study to have a good understanding of the possible implications of each approach in order to make the appropriate decision.

2.1.2 Identification of the Users and Purposes of the Evaluation Study

The users and purposes for an evaluation study heavily influence the type and amount of information to be collected, as well as the degree of precision required in its conduct. Evaluation studies which are carried out pursuant to the Treasury

Board Policy on Program Evaluation have as the client the deputy heads of departments and agencies. A deputy head may wish, that evaluation studies address in addition to his or her own concerns, the concerns of other parties such as the managers responsible for the operations of the program, non-governmental organizations that are instrumental to the delivery of the program, the Cabinet Policy Committees, and Treasury Board. In this sense, a study may have more than one user, but the deputy head has final responsibility for it.

Since different users may have different concerns about the program, and therefore different information needs, the type of evidence and the degree of precision they require to draw conclusions about the effectiveness of a program may vary. Although an evaluation study may not respond to all the information requirements of all its users because of the time and costs involved, an evaluation assessment study should identify the major potential users of an evaluation and describe their information needs. This will provide the deputy head with a better basis for defining the focus, scope and methodological requirements of the evaluation study.

2.1.3 Identification of Potential Evaluation Questions

The client and users of an evaluation study often express their information requirements in general terms which need further specification. The identification of specific questions which may be investigated in an evaluation study is an essential step of the evaluation assessment study. It guides evaluators at this stage in developing evaluation options for consideration by the client.

The identification of potential evaluation questions should ideally be a two-stage process. In the first stage, the evaluator would translate the concerns expressed by the various users of an evaluation study, as they relate to program effectiveness, into specific topics for investigation. He or she would also consider all the generic issues outlined in Table 2.2 and translate them into issues that are specific to the program being evaluated. For example, the evaluator will identify the specific program objectives, impacts and effects for which plausibility and achievement may be assessed, and the specific alternatives for program design and delivery which may be examined. Based on his or her analysis of the program and the information needs of the potential users of the evaluation study, the evaluator would then formulate a list of potential effectiveness evaluation questions. In the second stage of the process, the evaluator would discuss this list

of questions with the client and his or her program managers. The purpose of this discussion is to identify and assign priorities to the questions for which evaluation approaches will be developed.

In cases where access to the client is limited, the evaluator should assess the technical and practical feasibility of addressing all the specific program-related evaluation questions based on the generic questions outlined in Table 2.2. This will help ensure that evaluation options available to the client are not limited, a priori.

2.1.4 Determination of Evaluation Approaches

Once potential evaluation questions have been identified, the next step is to decide how to answer them. During an evaluation assessment study, the emphasis is on identifying possible evaluation approaches and on describing the nature of the evidence that will be used to answer each evaluation question. Detailed evaluation methods and techniques such as the formulation of questionnaires ordinarily are developed during the evaluation study.

Many approaches and techniques are available to carry out an evaluation study, ranging from a controlled experimental design to case studies. However, no one approach is appropriate to all cases and each has its advantages and disadvantages. The task of the evaluator is, therefore, to select a logically tight approach which is tailored to the information needs of the client and users, and at the same time takes into account factors such as the accessibility of information on the program and the time and resource constraints.

The determination of possible evaluation approaches involves the identification of the following:

- the evaluation studies which have been carried out for the same or similar programs and their implications for the current evaluation;
- the indicators or set of measures that represent the objectives, impacts and effects of a program and which will be used to assess its performance and outcomes; and
- the evaluation methodologies or set of methods and techniques which could be used to provide the needed information.

Table 2.2

BASIC PROGRAM EVALUATION ISSUES

Classes of Evaluation Issues	Basic Evaluation Questions
PROGRAM RATIONALE (Does the program make sense?)	<ul style="list-style-type: none">- To what extent are the objectives and mandate of the program still relevant?- Are the activities and outputs of the program consistent with its mandate and plausibly linked to the attainment of the objectives and the intended impacts and effects?
IMPACTS AND EFFECTS (What has happened as a result of the program?)	<ul style="list-style-type: none">- What impacts and effects, both intended and unintended, resulted from carrying out the program?- In what manner and to what extent does the program complement, duplicate, overlap or work at cross-purposes with other programs?
OBJECTIVES ACHIEVEMENT (Has the program achieved what was expected?)	<ul style="list-style-type: none">- In what manner and to what extent were appropriate program objectives achieved as a result of the program?
ALTERNATIVES (Are there better ways of achieving the results?)	<ul style="list-style-type: none">- Are there more cost-effective alternative programs which might achieve the objectives and intended impacts and effects?- Are there more cost-effective ways of delivering the existing program?

Related Evaluation Studies

Consideration should be given to identifying and reviewing evaluations carried out on similar but not necessarily identical programs. Such a review enables the evaluator to become familiar with evaluation approaches that have been previously used. It also allows the evaluator to foresee practical problems that are bound to occur during the evaluation study and, therefore, develop strategies for dealing with them.

Evaluation Indicators

The following principles should guide the evaluator in selecting measures for evaluating the performance and results of a program.

The indicators, whenever possible, should be objective.

Evaluation indicators should, as far as possible, measure what they purport to measure without being influenced by individual characteristics particular to the evaluators, the clientele served by the program (if they are to be contacted), and the program managers and staff. In cases where it is only possible to use more subjective indicators such as expert opinion or polls, sufficient features should be incorporated into the evaluation design to compensate for the probable bias introduced by such indicators, or at least to identify explicitly such bias. For instance, samples could include persons with different stakes in the program; interviews could make use of indirect questioning; and additional questions may be added to a questionnaire to identify respondents' biases or to balance other questions which are likely to elicit answers that will unduly favour one side of an issue.

The indicators should comprehensively reflect the intent of the condition or situation they are intended to assess.

Evaluation indicators should provide relevant and sufficient information on the performance and results of a program. To help ensure that all aspects of a particular program objective or impact and effect are adequately measured, several indicators should be used, unless it is not cost-effective to do so. Similarly, different indicators are needed if there are genuine differences in views among program managers concerning the structure of a program, or what the program should do, or what it actually does.

Evaluation Methodology

There are some generally accepted principles which are useful to take into account when proposing or selecting a particular evaluation methodology. These are:

Assessing the effectiveness of a program requires that the actual activities of the program be adequately documented.

It is often impossible to infer program effectiveness without describing the major program activities that are actually carried out. Parliament and Cabinet can mandate a department to carry out a specific program, but statements of mandates are generally permissive in terms of what a program may do. Documenting the actual program activities permits the evaluator to learn what the program is and whether there are gaps in the delivery or quality of services. Such documentation can also help in analyzing the activities, or set of activities, that are important in bringing about impacts and effects, and can help in obvious ways in interpreting results which suggest that a program has not been effective or that it is not relevant. Many approaches can be used to document the actual activities of a program. They include: observing directly the manner in which a program is actually implemented; reviewing program records, interviewing line managers as well as clientele served by the program; and talking to personnel in related programs or activities (including other departments).

Program evaluation requires that the contribution of a program to the achievement of a desired result be determined with some confidence.

This implies that while credible analysis is always required, program evaluation should not be viewed as a scientific exercise aimed at producing definitive answers but rather as a tool for providing believable and timely information which is appropriate for decision-making and management. Accordingly, the relationship between a program and the results achieved should be established with some confidence to provide an adequate basis for making decisions about the relative value of a program. Establishing such a relationship requires comparing the results achieved with information that can be used to infer what the results would have been had the program not existed. Many means exist for making such inferences, all of which are more relevant in some situations than others.

They include: time-series analyses to detect changes over time; the development of logical arguments to show what the situation could be without the program; a comparison before and after entering a program with like persons or units that did not enter the program; and the use of modelling techniques (e.g. econometric models) to test inferences on relationships.

The assessment of alternatives to the program or its delivery system requires that the probable results of each alternative be determined in a systematic manner and that comparisons between such alternatives be made in terms of a set of relevant, comprehensive and consistent criteria.

Typically, in the context of program evaluation, the analysis of alternatives is limited to common sense and rational analysis as opposed to the use of sophisticated quantitative methods of analysis. Yet it is important, whenever an analysis of alternatives is to be undertaken, to explicitly identify the nature of the alternatives to be considered in the evaluation study, the procedures to be used for estimating the expected results of each alternative, and the dimensions along which each alternative will be compared. In planning for an analysis of alternatives, the evaluator should be concerned with the viability of the alternatives to be investigated in the evaluation study and the dependability of the information on the expected results of each alternative. The evaluator should also ensure that the criteria to be used for comparing the various alternatives are consistent and relevant to the client and users of the evaluation study.

When a generalization is extended from a sample group to a general population, care should be taken that the sample group be composed of persons or other entities of the same class as that to which the findings will be generalized, and that the sample is either large enough or appropriately weighted to reflect the results of the program with adequate sensitivity.

Many methods can be used to select samples, some of which will be more appropriate to certain situations than others. They include the major forms of non-probability sampling (e.g. purposive samples and quota samples) and probability sampling(e.g. random samples, stratified samples and cluster samples).

Whatever the approach selected, care should be exercised to ensure that the samples will include enough cases that are relevant to the populations in which the client and the users of the evaluation study are interested.

The methods used to collect information should be reliable.

Information can be gathered from such sources as interviews, questionnaires, direct observation, and available records and documentation on the program. No information collection method is dependable in all cases. Statistical records are often incomplete, and a good deal is known about common sources of bias in direct observation of events, in question wording and in interviewer behavior. The choice of a particular method or combination of methods requires that the evaluator be familiar with their advantages and limitations in relation to the context in which they will be used. In addition, the evaluator should know the procedures that could be followed for increasing the reliability of proposed information collection methods. For example, several sources of information could be used, specialized personnel in Statistics Canada could be consulted in the design of questionnaires, and trained interviewers could be selected for conducting interviews.

Technical trade-offs made in the course of an evaluation study should take into account the information needs of the client and users.

In the evaluation of decentralized programs or programs that are composed of several projects, the evaluator might compromise, for example, on the amount of information collected about a program, or individual projects in one setting, in order to increase the number of settings or projects sampled. This could be done if it were more important to have an estimate of the incidence of a particular phenomenon than to get detailed information about the relationships between the activities of a program (or projects) and the results achieved. The evaluator's choice in such technical matters should be made in the light of the information needs of the client and the users of the evaluation study.

2.1.5 Estimation of Costs and Resources

Once alternative evaluation approaches have been identified, the requirements for each approach should, as a minimum, be specified in terms of the following:

- (i) composition of the evaluation team: the number of program personnel, internal evaluation personnel, outside consultants and others, along with their related skills and expertise;
- (ii) the skills and experience of internal personnel used in the study (e.g. evaluation personnel, program personnel, and other internal personnel) and the duration of their involvement in the study (person-months);
- (iii) fees for consultants and specialists, if they are required;
- (iv) non-salary costs beyond the above fees; and
- (v) the time required to carry out the study.

2.1.6 Determination of Evaluation Options

Evaluation options are discrete packages of specific evaluation questions which could be answered in the subsequent evaluation study, along with their related approaches, resources and time requirements. A considerable amount of art and skill is needed to package questions and evaluation approaches in order that the collection of evaluation options provide the client with an adequate range of issues which could be investigated in the evaluation study, an adequate range of approaches which could be used to do so and consequently, the range within which the costs of the evaluation study may vary. Of particular importance in the design of evaluation options are the economies of addressing several questions with one approach.

The process of developing options involves reviewing the preceding steps of the evaluation assessment study and determining which questions should be answered. Some of the factors which should be taken into account in the determination of evaluation options are:

- the degree of resolution of all the information needs of the client;
- the feasibility of addressing particular aspects of the program in a manner that would satisfy the client's needs and ensure the credibility of the findings;
- the possibility of answering specific evaluation questions by the time program-related decisions are to be made; and

- the cost-justification of each evaluation option; that is, the costs of answering a specific set of questions through particular techniques relative to the costs of the program itself and the likely benefits which will derive from the study.

2.1.7 Formulation of Recommendations

Recommendations are normally formulated at the end of an evaluation assessment study. Evaluators ordinarily are expected to recommend to the study client a specific evaluation option or options. In some cases, postponement of the study to a later date may be recommended. For instance, evaluation studies may be recommended to be delayed if too many divergent views are held of the structure of a program, or if the program is not mature enough to be effectively evaluated, or if the cost's of such an evaluation are not commensurate with its benefits. Where it is recommended that an evaluation study be rescheduled to a later date, an evaluation assessment study should typically identify:

- (i) the conditions which make an evaluation study inappropriate at this particular time;
- (ii) the steps necessary to enable the evaluation of the program at a future date; and
- (iii) the evaluation approaches which could be used in the future evaluation of the program.

In all cases, it is important not only that recommendations be based on the results of the analyses performed in the assessment study, but also that the reasons for making specific recommendations be adequately documented and reported to the client.

2.2 QUALITY CHARACTERISTICS OF AN EVALUATION ASSESSMENT STUDY

Evaluation assessment studies are usually conducted before a commitment to an evaluation study is made. To make better use of them in the decision-making process and, at the same time, control the evaluation expenditures, it is important that adequate evaluation assessment studies be carried out. An adequate evaluation assessment study:

- is comprehensive in scope and content,
- yields relevant and credible products, and
- is cost-justified.

Comprehensiveness

An evaluation assessment study should investigate the feasibility of answering, in the ensuing evaluation study, all the issues of concern to the deputy head. Where the deputy head has not directed the assessment of only a limited number of issues, an assessment study should consider the four general classes of evaluation issues outlined in Table 2.2 as potential topics to be investigated in the subsequent evaluation study. This should result in an appropriate list of evaluation questions specific to the individual program and covering each of the four basic classes of evaluation issues.

Furthermore, the assessment should develop evaluation options covering groups of evaluation questions and approaches. The collection of options should provide the deputy head with information about the range of issues which could be investigated in the subsequent evaluation study, the range of evaluation approaches which could be used to do so, and consequently, the range within which the costs of the evaluation study may fall.

Finally, an evaluation assessment study should be complete enough to produce the following:

- (i) an updated program component profile;
- (ii) a general assessment of the program structure which indicates whether or not the program appears to operate as intended, and whether or not its activities and expected results are plausibly interrelated;
- (iii) evaluation options; and
- (iv) recommendations as to the appropriate courses of action.

Relevance

An evaluation assessment study should provide, on time, the information needed by the client to make decisions in relation to the focus, scope and depth of the subsequent evaluation study. An evaluation assessment study should therefore identify all the important evaluation questions and provide adequate information

on all the evaluation options. In particular, each option should specify (i) the questions which could be answered in the evaluation study; (ii) the evaluation approaches and techniques which could be used; (iii) the associated estimated resources and time requirements; and (iv) the likely credibility of the evaluation findings.

Credibility

An evaluation assessment study should, first of all, exhibit an understanding of the program and its environment, and provide demonstrable evidence to indicate whether or not:

- the activities of the program are consistent with its mandate; and
- a plausible relationship exists between the activities of the program and its expected impacts and effects.

To enhance credibility, the study should identify the known evaluation issues regarding the program. It should also provide new insights into the program and its underlying rationale, thereby identifying other evaluation issues not previously thought of.

Credibility is also enhanced when the range of evaluation options presented covers the particular interests of the deputy head, and the discussion of these options gives evidence that the benefits, limitations and feasibility of the various evaluation methods proposed have been explored. For example, the study should indicate that:

- the suggested evaluation indicators represent in an adequate and impartial manner the objectives, impacts and effects of the program or the alternative situations to be assessed;
- the information which is required on the activities of the program and its possible results is accessible and potentially reliable; and
- the proposed methodologies, including the methods for data collection and sampling, are cost-justified and will provide useful and reliable information on the performance and results of the program.

Finally, for the recommendations to be credible, they should be based on the results of the analyses performed in the evaluation assessment study and should explicitly incorporate the reasons for making them.

Cost-justification

The resources required to carry out an evaluation assessment study will be related to the complexity of the program and to the number and scope of evaluation issues which need to be explored. Typically, the resources required vary from one to five person-months. For instance, an average of one to two person-months are usually needed to carry out assessments of programs that do not have many interrelationships with other government programs, are highly centralized, or are well-documented. Where there are complex inter-departmental or intergovernmental features, an evaluation assessment study may take longer.

CHAPTER 3

THE EVALUATION STUDY

An evaluation study is generally conducted after its terms of reference, including a detailed work plan, have been approved by the deputy head of a department or agency. Terms of reference are elaborated upon in Chapter 4 of this document.

This chapter provides an overview of the principles that would be useful to take into account in the conduct of an evaluation study. It also describes the characteristics of a good quality evaluation.

3.1 PRINCIPLES FOR THE CONDUCT OF AN EVALUATION STUDY

An evaluation study for which detailed terms of reference have been developed typically involves the following steps.

- 1 . the collection of the information required to answer the evaluation questions specified in the terms of reference;
2. the preparation of the information collected in a format suitable for its subsequent analysis;
3. the analysis of the information;
4. the formulation of conclusions which relate to the evaluation questions asked;
4. the formulation of recommendations for decision-making or action;
and
6. the communication of the evaluation findings and recommendations to the deputy head and the major users of the evaluation study (see Chapter 5).

In cases where detailed terms of reference have not been developed, the evaluators should first develop a specific work plan as described in Chapter 4.

The following presents some principles which would be useful to follow in the execution of an evaluation study for which detailed terms of reference have been prepared.

3.1.1 Collection of Information

In planning an evaluation study (as discussed in Chapter 2), an evaluator normally weighs the merits of several methods for collecting the desired information and recommends the methods to be used. In the course of the evaluation study, the evaluators develop or refine the information collection techniques that have been selected; for instance, questions to be asked in a questionnaire or in an interview are formulated. They proceed then with the collection of the required information.

To help ensure the integrity of the information collection techniques, and the relevance, accuracy and completeness of the information collected, effective quality control procedures should be implemented and maintained during the evaluation study. Although the nature of the mechanisms for quality control may vary with the particular circumstances of a study, some form of the following is considered good practice in the conduct of evaluation studies:

- pilot testing of information collection methods;
- using more than one source of information;
- monitoring the collection of information;
- editing the information collected; and
- implementing procedures for handling non-response and attrition should sampling be used.

Pilot Testing of Information Collection Methods

This involves using the methods selected for collecting information on a small sample, and performing preliminary analyses of the information obtained. Such a procedure is particularly important and advisable when an instrument such as a questionnaire is to be used for the first time. The purpose of pilot testing is to identify and resolve problems and deficiencies in the information collection methods or in the form and usefulness of the information it gathers.

During a pilot test, it is possible to determine if the data collectors and the respondents understand the instrument (and therefore whether further training of interviewers, for example, is necessary), and if there is a need to eliminate, add or clarify some questions, to redesign a rating scale, or to modify a procedure so that the information obtained will be objective, relevant, accurate and complete. In addition, preliminary analyses should be performed to ensure that the information

obtained is indeed sufficient to answer the evaluation questions.

Use of Several Sources of Information

Typically, no one source of information is totally dependable. For example, program records and files may be incomplete or inaccurate, and information obtained through survey methods may be biased by the data collectors or the respondents. To increase the dependability of the information collected, it is considered good practice to use several sources of information whenever possible. For instance, information on an industrial incentives program may be obtained from both program files and industry records, or by surveying program managers and industry officials in order to check or complement the information gathered from one particular source.

Monitoring

Unexpected events can, and frequently do, occur during an evaluation study. Often the evaluator is surprised to learn that the actual program does not involve the activities it was presumed to have had, or that the information cannot be collected from a particular source or can be collected from that source only. Such unexpected events can compromise an evaluation to different degrees. Therefore, consideration should be given to implementing a system for periodically monitoring the relevance and quality of the information gathered about the activities of the program and its possible results, and for modifying the method used to collect information should the type or quality of information be problematic.

Information Editing

Editing the information collected ensures that inadvertent errors do not become part of the information base. Typical procedures for editing information include range checks to ensure that all transcribed information falls within predetermined plausible bounds; checks on keypunching and typing to ensure that the desired information is properly transferred from source documents; and internal consistency checks such as the examination of a sample of the information collected to ensure that it is reasonable.

Procedures for Handling Non-Response and Attrition

Changes in the original composition of a sample are usually inevitable

during the course of an evaluation study. Individuals may drop out from the sample and others may provide incomplete information. Since these changes may seriously bias the study, there should be procedures for following up non-response and accommodating attrition. Similarly, there should be appropriate mechanisms to determine whether non-respondents differ from respondents in ways which may bias the evaluation findings.

3.1.2 Preparation of the Information

The information collected during an evaluation study should be organized and displayed in a format which will facilitate the process of analysis. To allow for the proper use and analysis of the information, it is considered good practice to:

- use appropriate methods to organize and record the information collected; and
- implement effective quality control procedures.

Organization of the Information

Many methods can be used to summarize and record information, including frequency distributions, categories and various forms of tabular presentations. It is important to select a method that is relevant to the nature of the questions to be answered in the study. For example, if categories are established to classify information, the elements that distinguish among the different categories will have to be related to the characteristics that are investigated in the evaluation study. It is also necessary that the method used to organize and record the information collected during the study be appropriate to the techniques employed for its analysis. For instance, information that is used in qualitative analyses may be organized in categories or descriptive summaries, whereas information which is to be used in computer analysis should be coded into symbols and recorded on specified formats.

Quality Control Procedures

The implementation of effective quality control procedures ensures that transcribed information is accurate; that original information and data files are labelled and documented well enough to permit other individuals to refer to them and use them properly at a later date; and that reasonable safeguards are in place to insure against unauthorized access to the information or its loss.

3.1.3 Analysis of the Information

Once the information has been collected and prepared, it must be analyzed so as to answer meaningfully the specific evaluation questions investigated.

The analysis of information usually involves some or all of the following types of issues:

- (i) describing the samples of persons and sites included in the study, specifying their particular characteristics in either qualitative or quantitative (statistical) terms;
- (ii) describing any other sources of information used, their relevance to the issues addressed, and their relative dependability;
- (iii) specifying the nature of the program activities, which might be in terms of the amount and quality of services or goods delivered;
- (iv) relating the quality and quantity of services or goods delivered to the characteristics of the samples, that is, establishing who gets what and, as far as possible, for what reasons;
- (v) estimating the effects of the program by contrasting program performance with information about what performance could have been in the absence of the program, or had the program operated under different conditions or sets of assumptions;
- (vi) identifying the reasons for any observed effects, or perhaps for the failure to obtain effects; and
- (vii) estimating the potential results of alternatives to the program or its delivery system;
- (viii) relating the costs associated with the program and its alternatives, if any, to their respective results; and
- (ix) generalizing the results obtained from the information collected and the analyses performed to the populations from which the samples were drawn and if appropriate, to persons and settings not included in the study.

A variety of analytic procedures is available. To safeguard against drawing unjustified conclusions from the information gathered, it is useful, whatever the procedures selected, to take into account the following principles:

The analytic procedures should be related to the nature of the information collected during the evaluation study.

Appropriate application of an analytic procedure should be based on consideration of the information collected. That is: only appropriate statistics should be used to describe the information (e.g. the mean should not be used when the median is required); the analyses undertaken should be based on all the relevant information collected including the significant factors which influence the results of a particular program; and the analyses should clearly differentiate between indicators of the desired results of the program and indicators of the factors that might influence such results.

The logic of each method of analysis should be made explicit.

Many methods of analysis can be used to infer program effectiveness. They include the development of logical arguments, the use of hard models and the utilization of both qualitative and quantitative methods of analysis. Whatever the method selected, the evaluator should clearly specify its underlying logic. If qualitative methods of analysis are used, the evaluator should identify the stages of development of a logical argument, document the assumptions it involves, and identify what constitutes reasonable evidence to support or refute such an argument. Similarly, if quantitative methods of analysis are used, the evaluator should specify the assumptions underlying the use of a particular technique and acknowledge any deviations from, or violations of, these assumptions. Flaws in this step are not only directly damaging to the evaluation results but can seriously impair the credibility of the evaluation effort.

The use of several methods of analysis is encouraged when there is no known or well-established analytic procedure.

In the developing field of program evaluation, different conclusions are often reached by using different analytic procedures. This is especially true when the assumptions involved in a method of analysis have not been

practically tested and validated. In such cases, it is good practice to use, whenever feasible, several methods of analysis so as to increase the credibility of the evaluation conclusions.

The reasons for any environmental change that is observed after a program has been introduced, or the failure to induce change, should be adequately accounted for.

Should a change in the program environment be observed after the program has been in existence for a reasonable period of time, the evaluator should rule out alternative explanations which may account for the occurrence of such a change. Many means exist to rule out rival explanations for a program's results. They range from the development of logical arguments to the use of quantitative techniques. Similarly, if no change is observed the evaluator should investigate all the relevant reasons. The no-change syndrome is often due to the short life of the program, the poor logic behind its conception, or the manner in which the program is implemented.

Exploratory analysis of the influence of other factors on the results of an evaluation study should be undertaken, whenever necessary and useful.

Evaluators should always be scanning the information which has been gathered in order to see if other relationships may exist than those so far perceived. They should also interact with the major users of the evaluation in order to identify any unaccounted factors which may be influencing the study results. The identification of such factors after an evaluation report is completed can seriously lower the credibility of an evaluation. That is not to say, however, that evaluators should get involved in exploratory research, but judgement should be exercised to determine the amount of exploratory analysis that is required during the course of an evaluation study.

The units of analysis should be appropriate to the way the information was collected and the types of conclusions to be drawn.

This can perhaps best be illustrated by an example. For instance, in a study which investigates the effect of a health promotion program on the smoking habits of individual Canadians, information collected at the level of the cities (e.g. sales volume of cigarettes per capita) is not sufficient to

draw conclusions about individual residents. The sales of cigarettes in a city may be explained by a number of factors that are not necessarily related to its residents. For example, individuals may buy their cigarettes in the cities to which they commute for shopping. If such factors cannot be adequately accounted for in the interpretation of the data, it would be more reliable to collect information on the individual smokers in the cities affected by the program.

Appropriate tests of significance should be performed whenever findings are generalized to the population from which the samples were drawn, and sensitivity analyses should be carried out whenever uncertainty exists.

In generalizing the findings of an evaluation study to the populations from which the samples were drawn, tests of significance should be applied to determine if each finding is significant or the result of chance variation in the information collected, and judgment should then be exercised to interpret the results of the tests with some confidence. In addition, sensitivity analyses should be performed, whenever uncertainty exists, to establish the range within which the value of a particular finding might vary. This is most important where factors exist which may have significant but unclear influence on the findings of an evaluation.

The generalization of the evaluation results to settings other than those included in a sample requires that a careful comparison be made between the settings examined in the study sample and these other settings in order to identify the type and degree of existing similarities and differences, and to determine their influence on the evaluation results.

Caution should be exercised in the generalization of the evaluation results to settings which have not been examined in the study since the success of a program or project depends, in part, on the setting in which it took place. However, if there is a requirement to make such generalizations, the evaluator could repeat the evaluation study in diverse settings, or carefully analyze the effect that distinctive characteristics of these new settings may have on the evaluation results.

3.1.4 Formulation of Conclusions

In the classical scientific method, conclusions follow quite straightforwardly from the evidence gathered: the evidence rigorously proves the conclusions. While this may be possible in rare cases where experimental evaluation methods are used, it is more likely that the evidence gathered by other evaluation methods will not answer a specific evaluation question conclusively. In such cases, conclusions are formulated by combining the best evidence gathered under the circumstances and sound judgment.

The credibility of less definitive conclusions can be enhanced by gathering different types of evidence relating to the same evaluation question. Even if any one piece of evidence may not permit firm inferences to be made, a collection of such inferences all suggesting the same conclusion can, in the aggregate, provide the client with enough evidence to make a management decision. In addition, the evaluator should ensure that the conclusions formed are fully discussed and reviewed with the major users of the evaluation study.

3.1.5 Formulation of Recommendations

Recommendations are typically formulated at the end of an evaluation study. They suggest to the client alternative courses of action to deal with the evaluation findings. Although recommendations are viewed as advice, and as a consequence do not form an integral part of the study report, they should take into account the results of the analyses of the information collected and the conclusions derived therein. In addition, as far as possible, alternate recommendations should be put forward, each depending on different assumptions inherent in the evaluation.

3.2 QUALITY CHARACTERISTICS OF AN EVALUATION STUDY

A good-quality evaluation study would be preceded by adequate terms of reference. This will ensure that the evaluation study addresses the questions relevant to the information needs of the client, in a manner that is acceptable to him or her. A good-quality evaluation study should:

- address relevant and significant questions;

- yield credible findings¹ and recommendations; and
- be cost-justified.

Relevance and Significance

An evaluation study should answer, in a timely manner, the questions identified in its terms of reference. Typically, these questions would be developed during the evaluation assessment and would be based on the generic program evaluation questions illustrated in Table 2.2. In an evaluation assessment, care is exercised to ensure that the questions to be answered in the evaluation study are relevant to the information needs of the client and that they will yield new and important information, that is, information which either goes beyond what is already apparent to the users of the study, or which confirms the conceptions they have of the program.

Credibility

The information collected and the evidence produced should be reliable and related to the questions addressed in the evaluation study. To ensure the reliability of the information and evidence gathered, credible evaluation approaches should be used. The credibility of an approach is determined by many factors, especially the adequacy of the indicators used to represent and measure the performance and results of the program being evaluated; the appropriateness of the evaluation methodologies used for inferring what performance could have been had the program not existed or operated under different assumptions; the suitability of the methods of sampling and information collection; and the sufficiency of the analyses performed to interpret the information collected and answer the evaluation questions set forth in the terms of reference for the study.

Credible conclusions come from having the conclusions based on relevant and reliable evidence and information; from clearly identifying where sound judgment has been used rather than hard data, thus stating what assumptions have been made in reaching the conclusions; and from demonstrating that important factors were not overlooked.

¹Findings encompass the evidence and information collected during the study, and the conclusions derived therein. Conclusions are the answers to the evaluation questions set forth in the terms of reference, with an indication of any qualifications to such conclusions that may be necessary.

Finally, the credibility of the recommendations is enhanced by making clear their link to the conclusions; by clearly indicating the assumptions used in their formulation; and by giving an indication to the client that the recommendations were made with a full appreciation of their implications.

Cost-justification

The costs of an evaluation study should be viewed in the light of the value of the information it provides (e.g. its relevance, significance, and credibility) as well as in relation to the costs and importance of the program evaluated.

CHAPTER 4

TERMS OF REFERENCE

There is a need to develop a specific plan of action before either the evaluation assessment study or the evaluation study begins. This plan of action, or terms of reference, represents a contract between the client and the evaluators. Its purpose is to ensure that each party understands the specific objectives of the study and the basic approaches to achieving them.

4.1 CONTENTS OF TERMS OF REFERENCE

Terms of reference for evaluation assessment studies and evaluation studies should, as a minimum, contain the following elements:

- (i) a statement of the purposes of the study. For example, the purposes of an evaluation assessment study are usually to develop an updated profile of a particular program and to determine a set of options for the evaluation study, whereas the purpose of an evaluation study is to answer an agreed set of specific questions. These questions should be expressed in terms which are particular to the program to be evaluated. In general, specificity in the formulation of the terms of reference for both studies will help to reduce the likelihood of misunderstanding during a study and will provide a useful foundation on which to base the related final report;
- (ii) a detailed work plan, further elaborated upon in section 4.2, which indicates:
 - how the purposes of the study are to be achieved: that is, an explanation of the specific tasks to be done and the evaluation approaches to be used;
 - who is going to do each task;
 - when each and all tasks are to be completed;
 - what reports are to be produced and at what time; and
 - who are the recipients of such reports;

- (iii) a description of the composition of the study team and their reporting relationships. This includes identifying the key staff members; specifying the responsibilities of the team leader; and delineating his or her reporting relationships to the director of program evaluation and the program evaluation committees, if in existence;
- (iv) a clear statement of the authority (if needed) to do the study;
- (v) a specification of the estimated resources to be committed to the study and its different parts; and
- (vi) an outline of the procedures for amending the study plan.

4.2 THE DETAILED WORK PLAN

This is the action plan for both the evaluation assessment study and the evaluation study. It is an integral part of the terms of reference since it describes how the expected products of each study will be achieved.

4.2.1 Work Plan for the Evaluation Assessment Study

Because of the general nature of an evaluation assessment, the work plan will indicate, in broad terms, the approach that will be used to carry out the study as well as identify its expected duration, the professionals involved in its conduct, and the recipients of the assessment report.

4.2.2 Work Plan for the Evaluation Study

The development of a specific work plan for an evaluation study is a critical element of the evaluation process. This action plan is required to ensure that the logistics of data or information collection and analysis are given adequate attention, and to provide a basis for quality control and cost control. In cases where no evaluation assessment study has been carried out, there is a need to deal first with all the methodology issues described in Chapter 2 of this document, and then to provide direction as to what information will be collected and how it will be analyzed.

The development of a work plan for an evaluation option selected by the client often requires evaluators, to undertake first a quick critical re-analysis of:

- the evaluation option selected by the client, particularly as it pertains to the adequacy of the evaluation approaches for addressing a package of questions;
- the evaluation approaches evolved in the evaluation assessment study, especially with respect to the desirability and feasibility of using specific indicators of program effectiveness, particular methodologies for inferring effectiveness, and specified procedures of sampling and information collection; and
- the practical constraints that were anticipated in the evaluation assessment study, and any strategies that were considered for dealing with them.

The evaluators will then develop a specific plan for the collection of data or information and its subsequent analysis. This work plan will include typically:

- (i) a description of the program structure, that is, its activities, outputs, impacts and effects, and their interrelationships;
- (ii) a description of the specific evaluation approaches that will be used to answer the evaluation questions, including the evaluation indicators, the nature of the information to be collected, the methodologies for inferring program effectiveness, the sampling procedures and sample sizes, and the methods of data or information collection and analysis;
- (iii) a description of the nature of the quality control procedures to be implemented during the course of the study;
- (iv) the identification of the professionals involved in the conduct of the evaluation study;
- (v) the identification of the expected duration of the study, including dates of important milestones; and
- (vi) an indication of the reports to be produced and their recipients.

It is expected that resources beyond those committed for an evaluation assessment study will be required to develop the terms of reference for the evaluation study, including the detailed work plan. The time and resources required for this task will vary, however, depending on the information provided by the evaluation assessment study.

4.3 AUTHORIZATION OF TERMS OF REFERENCE

The authorization of the terms of reference for both studies is the responsibility of the deputy head of a department or an agency. However, to make more efficient use of his or her time, he or she may delegate the responsibility for approving the terms of reference for evaluation assessment studies to the director of program evaluation.

4.4 DISTRIBUTION OF TERMS OF REFERENCE

The terms of reference for evaluation assessment studies and evaluation studies should be distributed to the departmental program evaluation committees if they exist, and to the officials responsible for the management of the program to be assessed or evaluated. Program evaluation committees will need the terms of reference to enable them to carry out properly their monitoring responsibilities. Similarly, program managers need to understand fully the specific purposes and scope of a study so as to participate to the fullest extent possible in its conduct.

CHAPTER 5

EVALUATION REPORTING

This chapter describes the characteristics of good quality evaluation assessment and evaluation reports. These characteristics deal with the form, content, debriefing, and distribution of these reports.

5.1 FORM

Written Reports

Both the evaluation assessment study and the evaluation study should culminate in a written report, to provide a permanent record of the analyses carried out and their associated results. A written report facilitates the communication of the results to responsible officials. It also facilitates subsequent follow-up work to determine whether appropriate action has been taken.

Self-contained Conclusions

The conclusions of each study should be clearly stated and self-contained in a section of the related report.

Segregation of the Body of the Report from Recommendations

Recommendations resulting from either an evaluation assessment study or an evaluation study should not form an integral part of the report. It is good practice to keep the analyses and results of a study separate from the recommendations, as these typically take into account other information on the program, in addition to the evaluation findings. Any recommendations prepared as part of a study should, be kept separate from the major results of that study. This will facilitate their communication to senior officials within and, when required, outside a department.

5.2 CONTENT

All evaluation assessment reports and evaluation reports should be credible and useful, that is, they should be:

- sufficient,
- accurate,
- objective,

- adequately documented,
- clear and concise, and
- timely.

Sufficiency

Both the evaluation assessment reports and evaluation reports should provide the client with sufficient information about the results of the study and the methods which were used for data collection and analysis. This will provide perspective against which the significance of the results can be judged, and hence enable the client and the readers of the report to reach their own conclusions.

Evaluation assessment reports provide the basis for the selection of appropriate terms of reference for the subsequent evaluation study, therefore they should, as a minimum, include all the elements described in Table 5.1. Evaluation assessment reports may or may not include the terms of reference for the subsequent evaluation study. In the latter case, the terms of reference for the evaluation study should be presented separately to the client for approval. Similarly, evaluation reports should contain all the elements described in Table 5.2.

Accuracy

Evaluation assessment reports and evaluation reports should ideally contain no errors of fact or logic. To help ensure the accuracy of the final report, procedural controls should be implemented. These controls include the verification of the report by the director of program evaluation to determine if the data and information collected support the study results, and the discussion of these results with the managers responsible for the program being assessed or evaluated, before the report is issued.

Objectivity

All the relevant results of evaluation assessment studies and evaluation studies should be presented in a balanced manner. To help ensure the objectivity of the reports, evaluators should:

- (i) report all results unaltered. An evaluator is not free to report only results which support his or her expectations (or those of the client), or to omit reporting conflicting results (such as from other studies);
- (ii) explicitly identify the influence of any assumptions which have been made, in information or logic, on the results of the study;

- (iii) sufficiently qualify the results to help readers avoid drawing unsubstantiated inferences. For example, in an evaluation study, for every finding involving a comparative statistical analysis, the statistical technique and levels of statistical and practical significance should be reported; and
- (iv) give equal emphasis in the report to all points of equal significance.

Adequate Documentation

Sufficient documentation about the study should be maintained to permit subsequent re-analysis of the information collected. To be accessible, information about the study as well as information gathered about the activities of the program and its results should be properly labelled and filed.

Clarity and Conciseness

The information presented in evaluation assessment reports and evaluation reports should be written with a minimum of ambiguity. Specialized technical language should be avoided. Where this is not possible, a glossary of terms should be appended to the text. Similarly, the use of clearly labelled bar charts and other graphic techniques is encouraged. Such techniques not only permit easy understanding but also enable the readers to undertake their own exploratory analysis.

In addition, to make more efficient use of the reader's time, reports should not be overloaded with detail. Detailed information and analyses should be included in technical appendices, and an executive summary should accompany the study report.

Timeliness

The findings and recommendations of evaluation assessment studies and evaluation studies should be reported to the deputy head promptly to help ensure their use in the decision-making process. Similarly, it is not good practice to perpetuate draft reports. A final report should be available within a reasonable period of time from the completion of the study (e.g. three months).

5.3 DEBRIEFING

The findings and recommendations of an evaluation study should be fully discussed with the major users (including the line

managers who have participated in the conduct of the study) before the report is issued. Comments should be obtained from these users and, where appropriate, incorporated into the report.

This practice allows the evaluator to verify the factual content of the report and to view the findings and recommendations of the study from the user's perspective. The evaluator should be well prepared to defend the evaluation report. However, where agreement cannot be reached, the points of disagreement should be stated in the report.

5.4 DISTRIBUTION OF REPORTS

Evaluation assessment reports and evaluation reports should be submitted to the client to enable him or her to consider the findings and recommendations of the studies in the decision-making process. Reports should also be distributed to the managers responsible for the operations of the program, to other officials who may be responsible for following-up actions taken on approved recommendations, and to program evaluation committees, if any.

Table 5.1

**A SUGGESTED OUTLINE FOR EVALUATION
ASSESSMENT REPORTS**

1. **An Executive Summary** which includes:
 - the objectives of the study and the approach used for its conduct;
 - the key findings of the study; and
 - the evaluation options.

2. **An Introduction** which indicates:
 - the questions addressed in the evaluation assessment study; and
 - the approach used to conduct the assessment and the major constraints, if any, under which the study has been carried out.

3. **An Updated Program Component Profile** (See Table 2.1) which describes:
 - the background of the component; and
 - the program component's structure or structures if more than one view is held of the nature of the program.

4. **A Summary of the Analyses Conducted** which includes:
 - information on whether the program operates as intended;
 - an assessment of the degree to which the activities of the program are plausibly linked to the attainment of its desired results, and the reasons underlying this assessment;
 - a summary of the major approaches used in the previous evaluations of the program or those of similar programs;
 - a presentation of the specific questions which could be answered in the subsequent evaluation study and the genesis of these questions; and
 - a presentation of the evaluation approaches that could be used to answer each evaluation question.

5. **A Presentation of Evaluation Options** with each option indicating:
 - a set of specific questions which could be answered in the subsequent evaluation study;
 - the related evaluation approaches which could be used and the reasons for their selection, including the identification of the evaluation indicators and methodologies;
 - the confidence with which each question could be answered; and
 - the time and resource requirements for dealing with that particular option.

Table 5.2

A SUGGESTED OUTLINE FOR EVALUATION REPORTS

1. **An Executive Summary** which includes:
 - the specific questions asked in the study;
 - the approaches used to answer them; and
 - the key findings of the study.

2. **An Introduction** which includes:
 - a program component profile (See Table 2.1);
 - the authority to evaluate the program and the specific questions investigated in the study; and
 - a statement of the major assumptions upon which the findings of the evaluation study are based, and a list of the major constraints, if any, under which it has been carried out.

3. **A Presentation of the Evaluation Approaches Used** - including the identification and description of, and whenever applicable, the justification for:
 - the indicators used to assess the effectiveness of the program; and
 - the evaluation methodologies used to collect the needed information and infer program effectiveness.

4. **A Summary of the Analyses Conducted** which identifies and describes:
 - the procedures implemented to ensure the reliability of the information collection methods, the accuracy of the information actually collected, and the appropriateness of the samples if sampling is used; and
 - the procedures used for the analysis of the information and their underlying logic.

5. **A Presentation of the Evaluation Findings** which includes:
 - a summary of the relevant information and evidence collected;
 - the conclusions of the study as they relate to each of the evaluation questions asked with an indication of any necessary qualifications to such conclusions; and
 - a statement of how dependable the evaluation findings are.

6. **Technical Appendices** which include detailed documentation on data collecting and analysis.