

BC Tribunal Dispute Resolution
Needs Assessment Project

- Report 3 -

**Building Tribunal Dispute Resolution Program
Evaluation Capacity**

by

Craig Darling

Accord Canada Dispute Resolution Group

March 2007

Contents

Introduction	1
Research Tasks.....	2
Study Sample.....	2
Research	4
Tribunal Survey	5
Interviews.....	5
Literature Search	7
Definitions and Fundamental Concepts	8
A Framework for Dispute Resolution Program Evaluation.....	13
Recommendations.....	16
Appendices	17
ADR Conference Performance Measures.....	19
The Program Evaluation Tool Kit.....	21

Introduction

In December 2005, the Administrative Justice Office (AJO) and Dispute Resolution Office (DRO) of the Justice Services Branch, Ministry of Attorney General, initiated a project to assess the need to enhance or expand dispute resolution¹ capacity in the BC administrative tribunal environment (the “needs assessment”).

The initial research² included a tribunal survey, which revealed that the participating tribunals do not see a need to expand or enhance their dispute resolution capacity.

Generally, the tribunals in the study sample indicated that:

- the users of their dispute resolution processes are satisfied, and
- tribunal dispute resolution programs are adequate.

These perceptions, together with tribunal reports of limited demand for enhanced or expanded alternative dispute resolution processes, suggest overall satisfaction with the status quo. However, this preliminary indication was qualified by three factors:

1. The project’s time, budget and resource limitations precluded direct interviews with tribunal chairs, making it difficult to determine the basis for the tribunals’ optimism.
2. The written survey primarily sought information about the ‘quantity’ not ‘quality’ of tribunal dispute resolution services, leaving some gaps in the baseline information.
3. Few tribunals evaluate their dispute resolution programs, and of those who do, most rely on ad hoc and informal methods with uncertain data collection and performance measures, making it difficult to objectively substantiate their positive ‘self-assessment’.

The preliminary needs assessment recommended that:

1. Gaps in the baseline information be addressed, through more research and if possible, in interviews, to permit a more fully informed assessment of the need to enhance or expand tribunal dispute resolution capacity.

¹ For the purposes of the needs assessment project and this report, dispute resolution process means “a process established by the tribunal to facilitate the settlement of one or more issues in dispute.” (Sec 1, [Administrative Tribunals Act](#), RSBC, 2004).

² See Darling, Craig R. *BC Tribunal Dispute Resolution Needs Assessment Project: Initial Research and Preliminary Assessment*. Administrative Justice Office and Dispute Resolution Office, Ministry of Attorney General, 2006.

2. More work on the nature and quality of evaluation is required to enable a meaningful assessment of tribunal dispute resolution processes.

A 'follow-up' project was designed by the AJO and DRO to implement these recommendations.

Research Tasks

The first goal of the follow-up project was to further inform the needs assessment by:

- verifying the results of the preliminary assessment with tribunal chairs, and
- completing further research to substantiate baseline data.

This work was undertaken by inviting participating tribunals to:

- review and comment on the accuracy of a brief summary of the tribunal's response to the phase one survey of tribunal dispute resolution capacity
- where necessary, provide additional information to fill gaps in the survey results, and
- comment on various elements of the capacity issue in interviews primarily designed to explore dispute resolution program evaluation.

The results of this research are described in a separate report.³

The second goal of the follow-up project was to begin work on a general organizational framework for dispute resolution program evaluation designed to enhance understanding and help tribunals identify the skills and resources required to conduct useful evaluation. This work was completed by searching the program evaluation literature, using criteria distilled from interviews with interested tribunals.

This report summarizes the results of the literature search.

Study Sample

The AJO and DRO identified eighteen BC administrative tribunals for the needs assessment study sample:

- Agricultural Land Commission (ALC)

³ See Darling, Craig R. *BC Tribunal Dispute Resolution Needs Assessment Project: Follow-up Research and Updated Assessment*. Administrative Justice Office and Dispute Resolution Office, Ministry of Attorney General, 2007.

- Community Care and Assisted Living Appeal Board (CCALB)
- Employment Standards Tribunal (EST)
- Environmental Appeal Board (EAB) / Forest Appeals Commission (FAC)
- Farm Industry Review Board (FIRB)
- Financial Services Tribunal (FST)
- Forest Practices Board (FPB)
- Hospital Appeal Board (HAB)
- Human Rights Tribunal (HRT)
- Labour Relations Board (LRB)
- Mediation and Arbitration Board (MAB)
- Mental Health Review Board (MHRB)
- Passenger Transportation Board (PTB)
- Property Assessment Appeal Board (PAAB)
- Safety Standards Appeal Board (SSAB)
- B.C. Utilities Commission (BCUC)
- Workers Compensation Appeal Tribunal (WCAT)

In February 2006, the AJO and DRO invited the chairs of the eighteen tribunals in the study sample to answer a brief questionnaire⁴ exploring:

- the nature and scope of the tribunal's dispute resolution processes
- current rates of utilization
- perceived barriers to utilization of dispute resolution processes
- perceived level of client satisfaction with and adequacy of tribunal dispute resolution processes
- the nature and extent of monitoring and evaluation of dispute resolution processes

Seventeen tribunals responded to the survey, fifteen of which confirmed that they use dispute resolution processes. The Financial Services Tribunal (FST) and the

⁴ See Darling, Craig R. [Initial Research and Preliminary Assessment](#), supra note 1, p. 6.

Mental Health Review Board (MHRB) reported that they do not use dispute resolution processes in their case management system.

In October 2006, the fifteen tribunals in the study sample who use dispute resolution processes were asked to verify the accuracy of a brief summary of the tribunal's response to the survey of tribunal dispute resolution capacity⁵, and, where necessary, provide additional information to fill gaps in the survey results. Thirteen tribunals replied,⁶ confirming that the summaries are accurate, and providing additional information, where possible.

The Property Assessment Appeal Board, Human Rights Tribunal, Forest Practices Board, Farm Industry Review Board, Environment Appeals Commission – Forest Appeals Commission, and Community Care and Assisted Living Appeal Board also agreed to participate in discussion, information sharing and consultation regarding the development of a dispute resolution program evaluation framework for administrative tribunals. Representatives of each of these tribunals were interviewed to:

- gain a better understanding of dispute resolution program evaluation practice in the tribunal community, and
- identify 'user-based' criteria to guide the search for a suitable framework for the evaluation of tribunal dispute resolution programs.

Research

The evaluation framework research included:

1. the initial survey of fifteen tribunals (to determine whether tribunals evaluate the effectiveness of their dispute resolution processes, and if so, what criteria are used)
2. interviews with representatives of seven tribunals interested in discussing dispute resolution program evaluation (to learn more about evaluation practice in the tribunal community and identify criteria for the literature search)
3. meetings with representatives of the DRO and AJO (to discuss government policy regarding tribunal dispute resolution programs)

⁵ See Darling, Craig R. [Initial Research and Preliminary Assessment](#), supra note 1, Appendix 1.

⁶ The Agricultural Land Commission and Labour Relations Board did not reply to the request.

4. a comprehensive literature search (to identify a program evaluation framework that might be suitable for adaptation by tribunals).

Tribunal Survey

The 2006 survey of tribunals revealed that:

1. Evaluation is not conducted consistently in the BC administrative tribunal environment; nor is it often well-integrated into case management procedures.⁷ A few tribunals have made evaluation an integral part of their dispute resolution programs, with verifiable performance measures, while others are much less systematic.
2. Among the tribunals who do evaluate their dispute resolution programs, data collection methods and performance measures vary. The common measures appear to be settlement rates, time savings and cost effectiveness. Only one of the tribunals in the study sample uses client satisfaction as a criterion for evaluating effectiveness.

Interviews

In November 2006, informal interviews were conducted with representatives of the seven participating tribunals.⁸ The content of the interviews varied depending on the nature and scope of the tribunal's evaluation work, but generally the purpose was to learn about each tribunal's evaluation goals and practices, discuss the need to expand or enhance evaluation work, and identify 'user-based' criteria to guide the literature search.

The interviews, together with the earlier survey, revealed that:

- There is broad notional support for the evaluation of dispute resolution programs. The potential value of program evaluation is acknowledged but implementation is not a priority. An often heard rationale is that evaluation is not a high priority because 'our resources are limited and we know intuitively that our dispute resolution program is working well'.

⁷ Only 2/3rds of participating tribunals (10 out of 15) evaluate their dispute resolution programs, and most of those rely on ad hoc and informal mechanisms not conducive to reliable results. See Darling, Craig R. [Initial Research and Preliminary Assessment](#), supra note 1, Table 7, p. 14.

⁸ Property Assessment Appeal Board, Human Rights Tribunal, Forest Practices Board, Farm Industry Review Board, Environment Appeals Commission – Forest Appeals Commission, and Community Care and Assisted Living Appeal Board

- Evaluation is often limited to occasional informal assessments, usually based on anecdotal or unsolicited data. With some exceptions, evaluation is generally not methodical – results lack precision and are hard to defend.
- There is no standard, consistent approach to dispute resolution program evaluation. This is explained as the natural consequence of diversity in the tribunal environment – different mandates, stages of development, size, case volume, personnel, training, ideology and perspectives on service delivery.
- There is a perception that:
 - program evaluation is demanding, expensive and hard to justify given other priorities
 - tribunal dispute resolution programs are not sufficiently well-developed to warrant sophisticated evaluation, and
 - the cost of evaluation may outweigh the benefits at this stage of dispute resolution program development.
- Program evaluation methodology is not widely understood.

These views are echoed in the program evaluation literature, which suggests that managers generally see evaluation as useful in theory but not in practice. Common sentiments include:

- concerns that evaluation often generates information that is either impractical or irrelevant — if the information is understood at all
- fears that traditional evaluation models are too big, too complex and too expensive, and
- sensitivities that evaluation is solely about performance — a test to prove the success or failure of a program (and its managers).⁹

The literature also suggests that the value of evaluation as a learning tool — “a reflective feedback mechanism for program development”¹⁰ is not widely acknowledged by program managers.

⁹ McNamara, Carter. *Basic Guide to Program Evaluation*, online at http://www.managementhelp.org/evaluatn/fnl_eval.htm#anchor1575679, adapted from the *Field Guide to NonProfit Program Design, Marketing and Evaluation*, Authenticity Consulting, LLC, online at http://www.authenticityconsulting.com:80/pubs/PG_gdes/PG_pubs.htm.

Discussion in the tribunal interviews suggested that program evaluation would likely be a higher priority among tribunal managers if the process was:

- straightforward (easy to use, potentially self-directed, or at least not requiring experts at every stage)
- understandable (free of complex jargon and processes, within the competency of managers not experienced in evaluation)
- manageable (small-scale, efficient and cost-effective)
- relevant (useful and appropriate)
- empowering (enabled better decisions), and
- adaptable (tailored to fit tribunal requirements).

These criteria were used to guide the literature search for a 'user friendly' evaluation framework, suitable for small-scale applications in the tribunal environment.

Literature Search

The search of the program evaluation literature was designed to identify a general organizational framework that would help tribunal managers:

- improve their knowledge of dispute resolution program evaluation, including the value of evaluation as a learning tool;
- identify the skills and resources required to conduct small-scale, in-house, meaningful process and outcome evaluation
- make decisions about their dispute programs, including what is working, what is not, and how their programs might be improved.

The literature search was completed in two phases. The first phase looked at program evaluation generally, to identify fundamental concepts and broad principles that might help guide evaluation in the tribunal context. The second phase focused on the identification of an evaluation framework capable of satisfying the needs of the tribunal community.

¹⁰ Administrative Conference of the United States (1995). Dispute Systems Design Working Group. *Evaluating ADR Programs: A Handbook for Federal Agencies*. Washington, D. C.: Administrative Conference of the United States.

Definitions and Fundamental Concepts

The research revealed a myriad of program evaluation methodologies, approaches and tools, some straightforward and others very sophisticated and difficult to comprehend. The following concepts and principles were distilled from the literature to help inform and guide the development of an appropriate approach to tribunal dispute resolution program evaluation.

What is program evaluation?

1. *Program evaluation is the systematic collection of information about the activities, characteristics, and outcomes of programs to make judgments about the program, improve program effectiveness, and/or inform decisions about future program development.*¹¹ It is a process for clarifying and measuring progress towards the achievement of program goals, providing a systematic means for determining what is working, what isn't – and what needs to be improved. The process may include qualitative and quantitative data, depending on the goal being measured.
2. Program evaluation is a means of informing decision-making in all stages of program development and implementation.
3. Program evaluation is conducted according to systematic, consistent and comprehensive guidelines aimed at ensuring the accuracy of the results.¹²
4. Evaluation doesn't have to be perfect to be a useful input for decision-making; it only has to be sufficient for the purpose. A high degree of statistical validity is a desirable but not always achievable goal. If context does not permit a high level of precision, evaluation can still be useful to the decision-making process (i.e., another tool in the process); taking limitations into account and bearing in mind that evaluation is seldom the sole basis for determining the future of a program. What this approach lacks in precision is made up by usefulness – the argument being that “20% of the effort generates 80% of the needed results.”¹³ The rationale is that it is better to do some evaluation than none at all.

¹¹ Patton MQ. *Utilization-focused evaluation: The new century text*. 3rd ed. Thousand Oaks, CA: Sage Publications, 1997.

¹² U.S. Department of Health and Human Services. Centers for Disease Control and Prevention. Office of the Director, Office of Strategy and Innovation. *Introduction to program evaluation for public health programs: A self-study guide*. Atlanta, GA: Centers for Disease Control and Prevention, 2005, p. 1.

¹³ McNamara, Carter. *Basic Guide to Program Evaluation*, supra note 9.

Types of Program Evaluation

5. There are many types of program evaluation, which can confuse and complicate evaluation planning.¹⁴ The preferred approach is to allow an appropriate evaluation design to flow from a determination of what the organization needs to know to make selected program decisions, rather than attempting to first understand and then choose from among the various types.¹⁵
6. The two key evaluations for tribunal dispute resolution programs are the process evaluation, which takes place early in the life of a program, and the outcome evaluation, which occurs at a time when a program is mature enough to permit a meaningful assessment of its impact.¹⁶
7. A process evaluation is conducted to improve the operation of an existing program. It is undertaken in the implementation phase to determine what is being achieved, whether implementation is consistent with the way the program was planned and, what improvements are required. Process evaluation is often characterized as a 'learning tool' — useful for emerging programs, enabling tribunals to chart progress towards goals.¹⁷
8. An outcome evaluation is conducted to determine whether a program is meeting its goals and is having the desired effect. Outcome evaluation is for mature tribunals, to assess effectiveness, and to determine results and overall value — often to help inform a decision about whether a program should be continued or expanded.¹⁸
9. In the tribunal context, a dispute resolution program evaluation would likely include elements of both process and outcome evaluation.

¹⁴ Types may be distinguished according to evaluative purpose (formative, summative, developmental), kind of data (quantitative, qualitative, mixed), kind of design (e.g., naturalistic, experimental), and kind of focus (processes, outcomes, impacts, costs, and cost-benefit, etc.). See Patton MQ. *Utilization-focused evaluation: The new century text*. 3rd ed. Thousand Oaks, CA: Sage Publications, 1997.

¹⁵ McNamara, Carter. *Basic Guide to Program Evaluation*, supra note 9.

¹⁶ Porteous, Nancy. *The Program Evaluation Tool Kit: a blueprint for public health management*. Ottawa-Carleton Health Department's Public Health Research, Education and Development Program and the Public Health Branch, Ontario Ministry of Health, 1997, p. 6.

¹⁷ Porteous, Nancy. *The Program Evaluation Tool Kit: a blueprint for public health management*, supra note 14.

¹⁸ Porteous, Nancy. *The Program Evaluation Tool Kit: a blueprint for public health management*, supra note 14.

Who should be involved and why?

10. Generally, program evaluation should include both people who have knowledge of the program and people with experience with evaluation.
11. Optimally, a program evaluation will include the persons or organizations with a stake in what will be learned from the evaluation and what will be done with the knowledge. This would generally include the program manager and staff, but also where practical:
 - those served or affected by the program (the primary users of the program including, for example, representatives of complainants disputants, advocacy groups, professional associations); and,
 - those in a position to do or decide something regarding the program (the primary users of the evaluation including, for example, tribunal chairs, ministry decision-makers and possibly senior government policy-makers).¹⁹
12. Canvassing a range of views and engaging a spectrum of interests early and throughout the evaluation process will help:
 - raise awareness and build support for the program and evaluation process
 - identify and clarify program goals and objectives (with immediate benefit in terms of shared understanding and relationship building)
 - determine the appropriate evaluation type
 - identify and integrate the tribunal's unique 'context' into the evaluation design, helping ensure an evaluation that is culturally appropriate.
 - focus the evaluation design on the key issues (potentially lowering the cost of evaluation)
 - determine the depth and breadth of information necessary to make informed decisions
 - enhance impartiality and credibility
 - ensure that the evaluation will provide the desired information and so enhance the usefulness of the results

¹⁹ *Introduction to program evaluation for public health programs: A self-study guide*, supra note 10, p. 11.

- facilitate the presentation and dissemination of results, and
 - improve the likelihood that results will be widely supported and actually implemented.
13. At a minimum, the primary users should be involved in identifying or clarifying program goals and identifying the questions the evaluation should focus on.
14. An evaluation specialist may be required at some stage in the process but not necessarily throughout.

When should evaluation be conducted?

15. Evaluation should be conducted throughout the life of a program, providing ongoing feedback to chart progress and make improvements. New programs should include an evaluation component from the outset. Emerging and established programs without an evaluation component should assemble baseline information in preparation for future evaluation and then integrate appropriate evaluation methodologies into the program.
16. Evaluation often occurs belatedly; for example, when hard evidence is required for funding or to otherwise justify the program. It is much more difficult to demonstrate that a program has made a difference without data to show how things were prior to implementation and a systematic evaluation that shows progress. Building evaluation into an existing program can be more expensive than designing a program with an evaluation component.

Why Evaluate?

17. The primary purpose of program evaluation is to enable managers to make better decisions. More specifically, evaluation can be used to ensure that dispute resolution services are being delivered effectively and efficiently, according to the preferences and needs of users by:
- monitoring progress toward achieving the program's goals
 - finding opportunities for improvement
 - providing the justification for ongoing funding and support, or
 - demonstrating compliance with government policy direction.
18. Program evaluation is generally driven by:

- an internal need to understand how well a program is working and how it might be improved, and/or
- an external requirement to demonstrate performance.

For example, in the tribunal community, evaluation might be useful to:

- tribunal managers who want to know how well their dispute resolution program is operating, so that they can learn from information gathered, and improve and expand it as required
- disputants and complainants who want to know that that their dispute or complaint is being resolved as quickly and fairly as possible
- government officials who want to know if tribunals' dispute resolution programs are delivering earlier solutions and faster justice
- government decision-makers who want to know the progress that is being made towards overarching government policy objectives.

How should programs be evaluated?

19. There is no single 'right' way to evaluate a program. The process must be tailored to the circumstances of the program, and the particular questions to be addressed. For example, tribunals are diverse and dispute resolution programs are context dependent – they must be evaluated in their individual contexts to produce meaningful reflection on their success.

20. An evaluation framework is a structured and systematic approach for assembling and organizing information about a program or some aspect of a program in order to make decisions about the program. A framework describes the broad, organizational steps in an evaluation process, which can be adapted and scaled as required to the needs of particular program.

21. An evaluation framework is a mechanism to help determine:

- What will be evaluated? (What is the "program" and in what context does it operate?)
- What aspects of the program will be considered when assessing performance?
- What standards (i.e., type or level of performance) can be used to indicate whether the program is successful?

- What evidence will be used to indicate how the program has performed?
- What conclusions regarding program performance can be made by comparing the available evidence to the selected standards?
- What lessons can be learned from the inquiry and how can those lessons be used to improve effectiveness?²⁰

A Framework for Dispute Resolution Program Evaluation

The second phase of the literature search sought to identify an evaluation framework that would be useful to the tribunal community, using the criteria distilled from the preliminary assessment and tribunal interview process. The research looked narrowly at alternative dispute resolution (ADR) evaluation processes, and then broadly at a range of program evaluation tools in a variety of fields.

In 1996, the US government passed legislation requiring the use of alternative dispute resolution in federal agencies. In anticipation of the legislation, the Dispute Systems Design Working Group of the Administrative Conference of the United States (the ADR Working Group) developed recommendations and performance measures to assist agency managers with implementation. The *Federal ADR Program Manager's Resource Manual*,²¹ which was the product of this work, describes a general framework and lists performance measures for ADR program evaluation in government agencies. The list offers a menu of measures from which agencies may select those appropriate to their evaluation needs (it is unlikely that all of the measures would apply to any single evaluation). The performance measures developed by the ADR Working Group appear throughout the evaluation literature, and have been used in government agency evaluation programs in Canada and elsewhere. The performance measures are listed in Appendix 1.

Aside from the work of the ADR Working Group, relatively little has been published about evaluation of ADR programs in government agencies and even less about ADR evaluation in the administrative justice system. Consequently, the literature search was expanded to include a general review of program evaluation in international conflict resolution, peacekeeping, public administration, community development, education, and public health.

²⁰ Centers for Disease Control and Prevention. *Framework for Program Evaluation in Public Health*. MMWR 1999; 48(No. RR-11), p. 4.

²¹ *Evaluating ADR Programs: A Handbook for Federal Agencies*, supra note 10.

The wider search revealed that the most advanced program evaluation work appears to be in the fields of education and public health. One of the most compelling examples is the *Framework for Program Evaluation in Public Health* developed and implemented by the US Centers for Disease Control and Prevention (the CDC framework).²² The CDC framework is intended to assist state, local and community managers and staff of public health programs in planning, designing, implementing and using the results of comprehensive evaluations in a practical way. It describes six broad steps that must be taken in any program evaluation and four sets of standards derived from the field of educational evaluation.²³

While the CDC framework focuses on the evaluation of public health programs, it could be applied in any program area, including tribunal dispute resolution programs. This raises the possibility of integrating the substance of the ADR Working Group recommendations with the process developed by the CDC, developing a new framework to support dispute resolution program evaluation in the BC administrative justice system. However, the comprehensive nature and scope of the CDC framework would make adaptation to smaller-scale initiatives challenging. Assessing the CDC framework against the 'user-based' criteria established in the tribunal interview process indicates the need for a less sophisticated approach.

Further research led to an innovative, capacity-building, evaluation initiative conducted in the Ontario public health system over a period of 4 years, from 1994 to 1997. This initiative resulted in the development of guiding principles for program evaluation, a self-directed learning resource called the *Program Evaluation Tool Kit: A Blueprint for Public Health Management* (the Tool Kit)²⁴, and an accompanying workshop. The project was motivated by the need to integrate evaluation into public health program management; notwithstanding the diversity of health units and the need to accommodate the concerns of disparate stakeholder groups.

²² Centers for Disease Control and Prevention. *Framework for Program Evaluation in Public Health*. MMWR 1999; 48 (No. RR-11)

²³ The standards for the CDC framework were derived from the seminal work of the Joint Committee on Standards for Educational Evaluation. See Joint Committee on Standards for Educational Evaluation. *Program evaluation standards: how to assess evaluations of educational programs*. 2nd ed. Thousand Oaks, CA: Sage Publications, 1994.

²⁴ *The Program Evaluation Tool Kit* is a practical, step-by-step guide to evaluating programs. It is presented in a series of short modules with simple explanations and specific tools for planning, conducting and using evaluation. The *Tool Kit* was developed by Nancy Porteous, Barbara Sheldrick and Paula Stewart in 1997 with funding from the Ottawa-Carleton Health Department's Public Health Research, Education and Development Program and the Public Health Branch, Ontario Ministry of Health.

The Tool Kit is a step-by-step guide to planning and conducting relatively small-scale evaluations, with an emphasis on process evaluation. An overview of the Tool Kit process is attached as Appendix 2 to this report. The complete Tool Kit may be viewed at <http://www.phac-aspc.gc.ca/php-ppsp/toolkit.html>. Although the resource is tailored specifically to public health, the evaluation process it presents is generic and can be applied in any program area.

The Tool Kit resource is written and presented at a level well-suited to the needs of small- to mid-size administrative tribunals. Indications are that it could satisfy the criteria distilled from the tribunal interviews, offering an evaluation framework that is

- straightforward (easy to use, potentially self-directed, or at least not requiring experts at every stage)
- understandable (free of complex jargon and processes, within the competency of managers not experienced in evaluation)
- manageable (small-scale, efficient and cost-effective)
- relevant (useful and appropriate)
- empowering (enabled better decisions), and
- adaptable (tailored to fit tribunal requirements).

While the Tool Kit could be used in its present form to guide the evaluation of tribunal dispute resolution programs, it would be more effective if adapted for tribunal use, with the menu of performance measures developed by the ADR Working Group as a resource. Adapting the Tool Kit to the tribunal context might, for example, include:

- incorporating dispute resolution (rather than public health) examples.
- amending the sample, worksheets and checklists to reflect the dispute resolution environment
- integrating existing dispute resolution performance measures (e.g., the ADR Working Group measures) into the data collection tool inventory as appropriate.

Recommendations

The second report in the BC Administrative Tribunal Dispute Resolution Needs Assessment Project²⁵ recommended that administrative tribunals, as part of their service planning, evaluate their dispute resolution programs to optimize service delivery in keeping with program goals and objectives, and overarching government policy. Towards that end, it is recommended that:

1. The AJO and DRO work with interested tribunals to explore the potential for applying the Program Evaluation Tool Kit to tribunal dispute resolution programs. Next steps might include:
 - circulating this report for comment, first among the tribunals who participated in the interview process, and subsequently among all tribunals in the study sample.
 - approaching the authors of the Tool Kit to:
 - i. discuss the application of the framework to dispute resolution program evaluation in the tribunal environment, and
 - ii. explore the prospect of adapting the Tool Kit resource to tribunal needs.
2. Subject to feedback received pursuant to recommendation 1, the AJO and DRO, in consultation with interested tribunals, adapt the Tool Kit to the tribunal context and test the evaluation framework in a case study or pilot program.

²⁵ See *Follow-up Research and Updated Assessment*, supra note 2.

Appendices

1. ADR Conference Performance Measures
2. The Program Evaluation Tool Kit

Appendix 1

ADR Conference Performance Measures

**PERFORMANCE INDICATORS FOR ADR PROGRAM
EVALUATION**

Dispute Systems Design Working Group

Administrative Conference of the United States

November, 1993

Table of Contents

Background Information 1

List of Indicators 2

Program Effectiveness

- Efficiency
 - Cost
 - Time
- Effectiveness
 - Dispute Outcomes
 - Durability of Dispute Outcomes
 - Impact on Dispute Environment
- Customer Satisfaction
 - Participants' Satisfaction with Process
 - Impact on Relationships Between Parties
 - Participants' Satisfaction with Outcomes

Program Design and Administration

- Program Organization
 - Program Structure and Process
 - Directives, Guides and Standards
 - Delineation of Responsibilities
 - Sufficiency of Staff
 - Coordination/Working Relationships
- Service Delivery
 - Access and Procedure
 - Case Selection Criteria
- Program Quality
 - Training
 - Neutrals

Other Specific Program Features

PERFORMANCE INDICATORS FOR ADR PROGRAM EVALUATION

Background Information

Evaluation is a key to determining whether an alternative dispute resolution (ADR) program has met or is successfully meeting its goals. It may also be used to assess the need for changes in the day-to-day administration of the program. Taking the time up front to carefully plan and design an evaluation of your ADR program will help to ensure that relevant information will be available to managers and decisionmakers to assess the effectiveness of the ADR program and to determine whether the program should be continued and/or modified.

This document is intended to serve several purposes. Ultimately, it will become part of a handbook on evaluation of Federal agency ADR programs. The handbook will address a whole range of issues that arise in the context of ADR program evaluation, including planning, designing, and implementing evaluations. In the meantime, this document is intended to provide initial guidance on the identification of both program goals and program measures. Program goals and measures are really two sides of the same coin. Goals are what your program seeks to accomplish; measures are used to determine whether those goals have been met.

The material contained in this document can be used in conjunction with the Administrative Conference's Dispute Systems Design Working Group's Pre-design Organizational Checklist to stimulate ideas about ADR program goals. It can also be used more directly to identify possible measures of success for ADR programs. It can therefore be used at both the "front" and "back" ends of program planning and implementation (design and evaluation, respectively).

Evaluations are conducted for different reasons and take different forms. Evaluation may be aimed at (1) determining whether the outcomes of a program are consistent with the program's declared goals, (2) determining whether the program is running the way it was intended to, and/or (3) determining whether changes in the program would improve its usefulness. Evaluations may be comprehensive in nature, rely on a significant degree of internal or external professional evaluation expertise, involve a great deal of planning, and take a rather lengthy time to complete. At the other end of the spectrum, evaluations may be aimed at providing more of a "snapshot" of where a program is at, at examining a particular area within a program, or at capturing the impact of specific changes in program coverage or administration. They may involve less planning and outside evaluation expertise, and take a relatively short period of time to complete. Or, the nature and form of an evaluation may fall somewhere in between these two ends of the spectrum. The reasons for which an evaluation is conducted, and the form it takes, will vary from agency to agency, and from time to time, depending on evaluation needs and constraints (e.g. budgetary), and each agency's particular mission/culture. Evaluations need to be designed to be responsive to managers and decisionmakers with different needs and interests.

The list of indicators below is divided into two categories, one dealing with **program effectiveness** (i.e. whether or not a program is meeting its goals), and one dealing with **program design and administration** (i.e. whether or not a program is being administered as it should be). (NOTE: the terms program measures and performance indicators are used interchangeably throughout this document to refer to specific ways of examining program effectiveness or administration.) These categories are not mutually exclusive, and the list itself is intended to be as comprehensive as possible, in order to cover to a wide range of agency interests/needs. It is unlikely that all of the measures listed below would apply to any single evaluation; rather, some will apply in some cases and others, not. Each measure is followed by one or more questions intended to further illustrate the kinds of evaluation issues an agency may wish to pursue.

The list, overall, is intended simply as a "sampling" of measures or indicators from which agencies may pick and choose, as appropriate, as they seek to formulate ADR program goals and to identify possible measures of program effectiveness.

List of Indicators

I. Program Effectiveness (Impact)

Program effectiveness measures or indicators are aimed at assessing the degree to which an ADR program is meeting its goals. More specifically, program effectiveness measures are used to examine the impact of the program on users/participants, overall mission accomplishment, etc. In the case of ADR, an agency may, for example, be interested in looking at whether the use of ADR reduces the time it takes to resolve cases/disputes.

Effectiveness indicators should correspond directly to the goals or objectives of an ADR program. For example, if a goal of your agency's ADR program is to reduce the backlog of cases, then the impact of the program on case disposition time needs to be assessed.

The indicators in the effectiveness category are further divided into three subcategories: **efficiency**, **effectiveness**, and **customer satisfaction**.

A. Efficiency

1. Cost

Cost to the Government of using ADR vs. traditional dispute resolution processes (e.g. negotiated settlements, agency findings, litigation).

Is the use of ADR more or less costly than the use of traditional means of dispute resolution? (Cost may be measured in staff time, dollars, or other quantifiable factors.)

Cost to disputants of using ADR vs. traditional dispute resolution processes.

Is the use of ADR more or less costly than the use of traditional means of dispute resolution? (Cost may be measured in terms of staff time, dollars, or other quantifiable factors.)

2. Time

Time required to resolve disputes using ADR vs. traditional means of dispute resolution.

Are disputes resolved more or less quickly using ADR processes, compared to traditional means of dispute resolution? Such factors as administrative case processing, participant preparation, dispute resolution activity timeframes, and/or days to resolution may be considered.

B. Effectiveness

1. Dispute Outcomes

Number of settlements achieved through the use of ADR vs. traditional dispute resolution processes.

Does the use of ADR result in a greater/fewer number of settlements?

Number of cases going beyond ADR steps.

Does the use of ADR result in a greater/fewer number of investigations, further litigation activities, etc.?

Nature of outcomes.

What impact does the use of ADR have on the nature of outcomes, e.g. do settlement agreements "look different," as in terms of the agreement or monetary amounts agreed upon? Do settlement agreements reflect more "creative" solutions?

Do outcomes vary according to the type of ADR process used?

Relationship, for cases selected for ADR, between dispute outcomes and such factors as complexity or number of issues, or number of parties.

Is there any relationship, where ADR is used, between the complexity and/or number of parties/issues in a case and the outcome of the case?

2. Durability of Outcomes

Rate of compliance with settlement agreements.

Does the use of ADR result in greater/lesser levels of compliance with settlement agreements?

Rate of dispute recurrence.

Does the use of ADR result in greater/lesser levels of dispute recurrence, i.e. recurrence of disputes among the same parties?

Impact on program/organizational environment.

Does use of ADR have the effect of improving the work environment, e.g. reducing the level of conflict and improving participant relationships--thereby contributing positively to mission accomplishment?

3. Impact on Dispute Environment

Size of case inventory.

Does the use of ADR result in an increase/decrease in case inventory?

Types of disputes.

Does the use of ADR have an impact on the types of disputes that arise?

Negative impacts.

Does the use of ADR have any negative consequences, e.g. an inability to diagnose and correct systemic problems/issues?

Timing of dispute resolution.

Does the use of ADR affect the stage at which disputes are resolved?

Level at which disputes are resolved.

Does the use of ADR have any impact on where and by whom disputes are resolved?

Management perceptions.

What are the quantitative and qualitative effects of using ADR on management, e.g. how does the use of ADR impact upon allocation and use of management time and resources? Does the use of ADR ease the job of managing?

Public perceptions.

Is the public satisfied with ADR outcomes? Is there any perceived impact of use of ADR on effectiveness of the underlying program? (NOTE: "Public" may be defined differently, depending on the particular program/setting involved.)

C. Customer Satisfaction

1. Participants' Satisfaction with Process

Participants' perceptions of fairness.

What are participant perceptions of access to ADR, procedural fairness, fair treatment of parties by neutrals, etc.?

Participants' perceptions of appropriateness.

What are participant perceptions of appropriateness of matching decisions (i.e. matching of particular ADR processes to particular kinds of disputes or specific cases)?

Participants' perceptions of usefulness.

What are participant perceptions of the usefulness of ADR in the generation of settlement options, the quantity and reliability of information exchanged, etc.?

Participants' perceptions of control over their own decisions/"destiny."

Do participants' feel a greater or lesser degree of control over dispute resolution process and outcome through the use of ADR? Is greater control desirable?

2. Impact on Relationships Between Parties

Nature of relationships among the parties.

Does the use of ADR improve or otherwise change the parties' perceptions of one other? Is there a decrease/increase in the level of conflict between the parties? Are the parties more or less likely to devise ways of dealing with future disputes? Are the parties able to communicate more directly/effectively at the conclusion of the ADR process and/or when new problems arise?

3. Participants' Satisfaction with Outcomes

Participants' satisfaction with outcomes.

Are participants satisfied/unsatisfied with the outcomes of cases in which ADR has been used?

Participants' willingness to use ADR in future.

Would participants elect to use ADR in a future dispute(s)?

II. Program Design and Administration (Structure and Process)

How a program is implemented will have an impact on how effective a program is in meeting its overall goals. Program design and administration measures or indicators are used to examine this relationship and to determine how a program can be improved.

The indicators in the program design and administration category are further subdivided into three subcategories: **program organization**, **service delivery**, and **program quality**.

A. Program Organization

1. Program structure and process.

Are program structure and process consistent with underlying laws, regulations, executive orders, and/or agency guidance?

Do program structure and process adequately reflect program design? Are program structure and process adequate to permit appropriate access to and use of the program?

2. Directives, guides, and standards.

Do program directives, guides, and standards provide staff/users with sufficient information to appropriately administer/use the program?

3. Delineation of responsibilities.

Does the delineation of staff/user responsibilities reflect program design? Is the delineation of responsibilities such that it fosters smooth and effective program operation?

4. Sufficiency of staff (number/type).

Is the number/type of program staff consistent with program design and operational needs?

5. Coordination/working relationships.

Is needed coordination with other relevant internal and external individuals and organizations taking place? Have effective working relationships been established to carry out program objectives?

B. Service Delivery

1. Access and Procedure.

Participant access to ADR option.

Are potential participants made aware of the ADR program? Is the program made available to those interested in using ADR?

Relationship between participant perceptions of access and usage of ADR.

What impact does participants' perceptions about the availability of the program have on the levels of program usage?

Participant understanding of procedural requirements.

Do program users understand how the program works? Did they feel in advance that they were comfortable with the process?

Relationship between procedural understanding and rates of usage.

Is there any relationship between the level of participant understanding and the degree of program use, e.g. is a lack of participant understanding serving as a disincentive to using the ADR program?

2. Case Selection Criteria.

Participant perceptions of fairness, appropriateness.

Do participants feel that appropriate types of cases are being handled in the ADR program? Do participants or non-participants feel that the criteria for which cases are eligible for ADR are fair? Are cases being sent to the ADR program at the appropriate dispute stages?

Relationship between dispute outcomes and categories of cases.

Is there a correlation between the [nature, size, types of disputants, and/or stage of the dispute] of cases and the outcome of the dispute? Are certain types of cases more likely to be resolved through ADR than others?

C. Program Quality

1. Training.

Participant perceptions of the appropriateness of staff and user training.

Do participants feel that they were provided with sufficient initial information and/or training on how to use the ADR program/process? Do they feel that program staff had sufficient training and/or knowledge to appropriately conduct the ADR program?

Relationship between training variables and dispute outcomes.

Is there a relationship between the type/amount of training (for participants and/or staff) and dispute outcomes?

2. Neutrals.

Participant views of the selection process.

Are participants satisfied with the manner in which neutrals were selected and assigned to cases? Were they involved in the selection decision? If not, did they feel they should be?

Relationship between participant views of the selection process, perceptions of neutral competence and objectivity, and dispute outcomes.

Is there any relationship between participant views about the selection process and dispute outcomes? How do these views affect participants' assessment of the competence and neutrality of neutrals?

Participant perceptions of competence (including appropriateness of skill levels/training).

Do participants feel that neutrals were sufficiently competent? Do participants feel that neutrals were sufficiently well-trained? Do participants feel that more or less training was needed?

Participant perceptions of neutrality/objectivity.

Do participants feel that neutrals were sufficiently objective? Do participants feel that neutrals were fair in their handling of the dispute?

Relationship between perceptions of neutral competence and neutrality and dispute outcomes.

Do participants' perceptions of the skills and/or objectivity of neutrals have any impact on the outcome of the dispute?

D. Other Specific Program Features

Every ADR program is unique. Those requesting and/or conducting an evaluation may want to consider examining other aspects of the ADR program. These unique features

may relate to the design of a program, who was and continues to be involved in program design and administration, etc. Each is likely to have at least some impact on service delivery and the quality of the program and should be considered for inclusion in either a comprehensive or selected evaluation of the program, as appropriate.

This document was developed by the Evaluation Subgroup of the Dispute Systems Design Working Group, Administrative Conference of the United States (ACUS).

Deborah Lesser, Department of Health and Human Services (Co-Chair)

Nancy Miller, ACUS (Co-Chair)

Peter Bloch, Nuclear Regulatory Commission

Dick Coccozza, Western States Senate Coalition

Anne Larkin, Department of Defense (OSD/WHS)

Leslie Van der Wal, Internal Revenue Service

Appendix 2

The Program Evaluation Tool Kit

Overview: The Program Evaluation Tool Kit

An innovative, capacity-building, evaluation initiative conducted in the Ontario public health system over a period of 4 years, from 1994 to 1997, resulted in the development of guiding principles for program evaluation, a self-directed learning resource called the *Program Evaluation Tool Kit: A Blueprint for Public Health Management*¹, and an accompanying workshop. The initiative was motivated by the need to integrate evaluation into program management; notwithstanding the diversity of public health units and the need to accommodate the concerns of disparate stakeholder groups.

Guiding Principles

The Guiding Principles for Evaluation in Ontario Health Units outline when, how, and why evaluations should be conducted and who should be involved, stressing:

- integration of program planning and evaluation
- the necessity for a clear description of the program
- the importance of tying the purpose of the evaluation to specific decision-making needs
- the need to ask specific evaluation questions
- ethical conduct; use of systematic methods
- clear and accurate reporting; timely and widespread dissemination
- a multidisciplinary team approach
- stakeholder involvement, and
- utilization of evaluation findings.

¹ The Program Evaluation Tool Kit is a practical, step-by-step guide to evaluating programs. It is presented in a series of short modules with simple explanations and specific tools for planning, conducting and using evaluation. The Tool Kit was developed by Nancy Porteous, Barbara Sheldrick and Paula Stewart in 1997 with funding from the Ottawa-Carleton Health Department's Public Health Research, Education and Development Program and the Public Health Branch, Ontario Ministry of Health. The Tool Kit may be viewed at <http://www.phac-aspc.gc.ca/php-ppsp/toolkit.html>.

Guiding Principles for Program Evaluation

Definitions

A **principle** is defined as a general law which guides action.

A **program** is defined as a series of activities supported by a group of resources intended to achieve specific outcomes among particular target groups.

Program evaluation is the systematic collection, analysis and reporting of information about a program to assist in decision-making.

Stakeholders are individuals and groups (both internal and external) who have an interest in the evaluation, that is, they are involved in or affected by the evaluation. Stakeholders may include program staff or volunteers, program participants, other community members, decision-makers, and funding agencies.

Guiding Principles

Integrated Program Planning and Evaluation

- Evaluation should be an integral part of program management and should occur during all phases of a program.
- All program plans should include how and when programs will be evaluated.

Clear Description of the Program

- The program being evaluated should be clearly described, especially process and outcome objectives, as well as intended target groups. Program logic models should be used when appropriate.
- Program objectives that are not specific should be clarified before continuing with further evaluation activity.

Explicit Purpose for Identified Need

- The purpose of any evaluation should be explicit and based on identified decision-making needs.

Specific Evaluation Questions

- Evaluation questions should be specific and clear.
- Evaluation questions should be based on the need to answer key management questions.
- The developmental stage of a program, its complexity and the reason for evaluating should be considered in formulating evaluation questions.
- Evaluation questions directly reflect a program's process and/or outcome objectives.

Ethical Conduct

- Members of the evaluation team should consider the ethical implications of program evaluation to ensure the rights of participants in the evaluation are respected and protected.

Systematic Methods

- The evaluation questions should drive the evaluation methods utilized.
- A review of the literature and a scan of evaluation activity in relevant program areas in other health units should be carried out at the outset of the evaluation.
- New data should not be collected if existing information can adequately answer evaluation questions.
- The most rigorous evaluation methods should be used given time and resource limitations.
- Evaluation should employ information (quantitative or qualitative or both) gathered from a variety of sources with varying perspectives.

Clear and Accurate Reporting

- Evaluation reports should include a description of the program and its context, the purpose of the evaluation, information sources, methods of data analysis, findings and limitations.
- Evaluation reports should be presented in a clear, complete, accurate, and objective manner.

Timely and Widespread Dissemination

- The dissemination of evaluation findings to stakeholders should be timely.
- Evaluation findings should be shared with other Ontario health units when appropriate.

Multidisciplinary Team Approach

- The evaluation team should include a variety of people who have adequate knowledge of the program, its participants, and program evaluation.
- Responsibilities should be agreed upon at the beginning of the evaluation. One person should be responsible for the overall management of the evaluation.
- The evaluation team should seek technical advice, support, and/or training, when necessary.
- Members of the evaluation team should continuously work toward improving their program evaluation skills; team members with evaluation expertise should support this learning.

Stakeholder Involvement

- Stakeholders should be consulted and, if appropriate, involved directly, throughout the evaluation process, within time and resource limitations.
- Stakeholders' interests, expectations, priorities, and commitment to involvement should be assessed at the outset of the evaluation.
- Communication among stakeholders should be honest and open.
- Evaluation should be sensitive to the social and cultural environment of the program and its stakeholders.

Utilization of Evaluation Findings

- Program managers should formulate an action plan in response to evaluation findings.
- Evaluation findings should be used to support decision-making.

The Program Evaluation Tool Kit

The Program Evaluation Tool Kit is a step-by-step guide to planning and conducting relatively small-scale evaluations, with an emphasis on process evaluation. Although the resource is tailored specifically to public health, the evaluation process it presents is generic and can be applied in any program area.

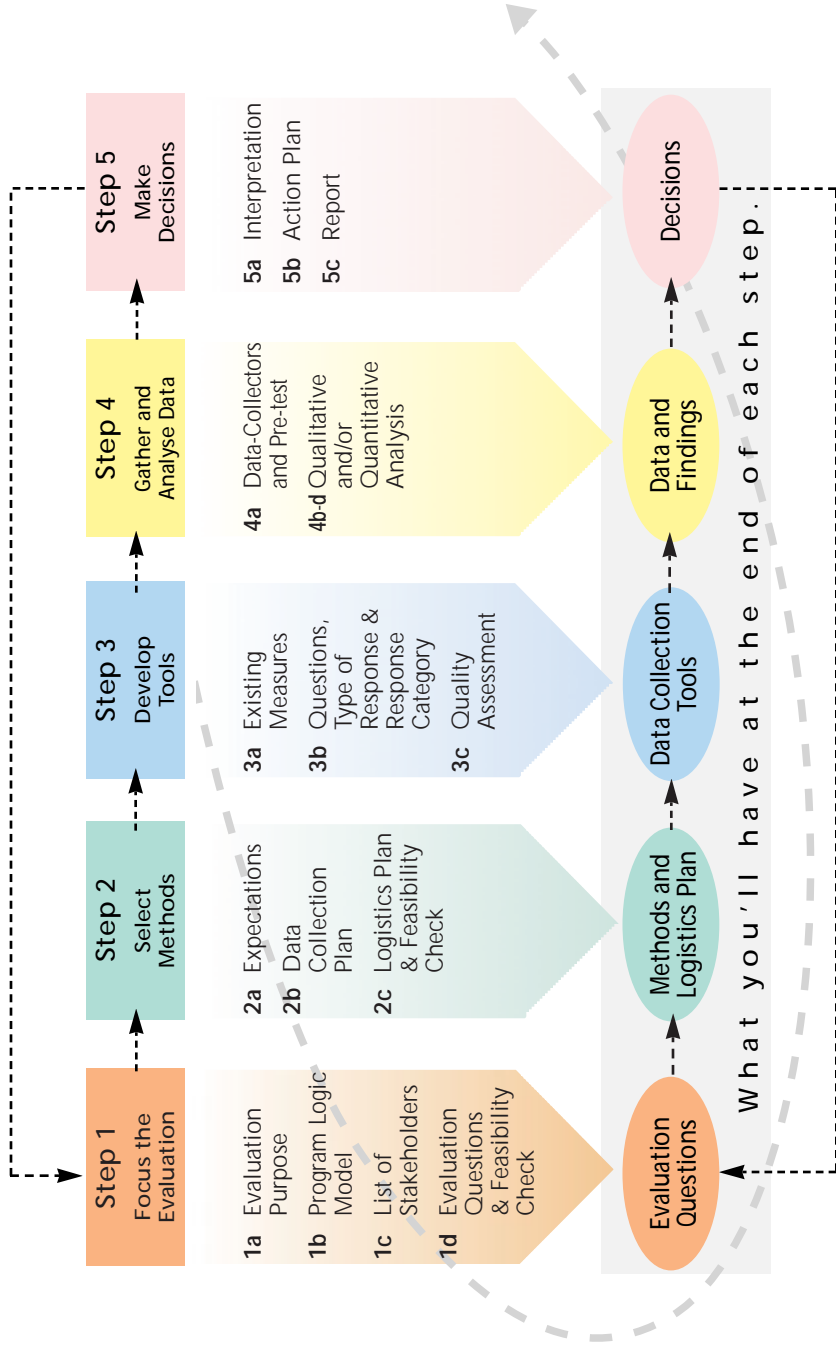
The Tool Kit defines program evaluation as “the systematic gathering, analysis, and reporting of data about a program to assist in decision making.” This definition emphasizes that:

- evaluation is a systematic, step-by-step process
- questions about any aspect of a program can be answered, depending on the stage of the program’s development, and
- evaluation serves specific program management decision-making needs.

The Tool Kit is designed to:

- improve managers’ knowledge and skills in the evaluation process to help integrate evaluation into program management (whether the evaluation work is conducted in-house or by an external expert)
- help managers make better decisions about their programs — decisions about ways to improve programs and decisions about the best use of program resources
- structure evaluation as a step-by-step process that can address any aspect of a program, depending on its stage of development; and
- focus on the evaluation process (not specific indicators or data collection instruments).

The Tool Kit distills the evaluation process to five steps, integrating planning and evaluation and emphasizing the ongoing, cyclical nature of the process. The steps are illustrated in the diagram on page 5.



The ongoing, cyclical nature of evaluation is illustrated in Figure 1, a diagram of the Tool Kit's step-by-step approach. Step 1 explains how to focus the evaluation by clarifying the purpose of the evaluation, describing the program, consulting stakeholders, and drafting evaluation questions. Steps 2 through 4 deal with collecting and analyzing data to address the evaluation questions. Step 5 draws on the findings of the evaluation to make decisions about ways to improve the program. An action plan for implementing changes to the program is a key activity in Step 5.

Step 1- Focus the Evaluation

The first step in evaluating any program is deciding the focus of the evaluation. This step lays the foundation for the remaining steps in the evaluation process. Step 1 of the Tool Kit will help determine:

- Why the program is being evaluated at this particular point in time?
- How the program is supposed to work, with whom, and why?
- Who will be using the evaluation findings?
- Precisely what information is needed from the evaluation?

Identifying the Purpose of the Evaluation

It is important to identify the broad purpose of the evaluation before addressing how the evaluation will be conducted. The key question is "why is the program being evaluated now?" For example, the purpose of the evaluation might be to:

- identify strengths and weaknesses
- share experiences
- measure progress
- improve implementation
- see what has been achieved so far, or
- make decisions about which programs or aspects of programs should be continued or discontinued.

Creating a Logic Model

There are five basic elements in any program: components; activities; target groups; short-term outcomes; and long-term outcomes.

A program logic model is a tool for clearly describing the program, showing what (components, activities) the program is supposed to do, with whom (target groups) and why (long and short-term outcomes).

For evaluation purposes, a logic model can:

- summarize the key elements of the program (hopefully on a single piece of paper)
- explain the rationale behind program activities
- show the cause-and-effect relationships between the activities and the outcomes -- that is, which activities are expected to lead to which outcomes
- help identify the critical questions for the evaluation, and
- provide the opportunity for stakeholders in the evaluation to discuss the program and agree upon its description.

Logic models are also a useful means of communicating the elements of the program to policy makers, staff, external funding agencies, the media and colleagues.

Consulting with Stakeholders

The next step is to consult stakeholders – the individuals or groups who have an interest in the program’s evaluation. Stakeholders may be internal or external to the organization.

Stakeholders are identified by asking “Who will be using the information from the evaluation?” Bear in mind that, for the evaluation to be credible and useful, it should be focused on serving the information needs of a few primary users

Examples:

- program managers

- staff
- ministry officials
- program participants
- professional associations

Determining the Evaluation Questions

The next step is to determine the questions the evaluation will seek to answer.

The evaluation cannot be all things to all people. It is crucial to limit the evaluation questions to high-priority issues only. Evaluation questions range from broad to specific depending on the stage of the process.

STEP	EXAMPLE
Purpose of the evaluation (Step 1a)	How can this program be improved? Is this program worth continuing?
Evaluation question (Step 1d)	Have the program's short-term outcomes been achieved?
Question on a questionnaire (Step 3)	On a scale from 1 to 5, where 1 is not at all useful and 5 is extremely useful, how would you rate the usefulness of the process you participated in?

The evaluation questions depend on decision-making needs and flow directly out of the program logic model. There are many commonly asked evaluation questions but there may also be some unique to the program being evaluated.

Step 2 – Select Methods

The second step in any evaluation is to determine the best way to answer the evaluation questions by zeroing in on the specific data required and the methods that should be used to gather them. This involves establishing the expectations for the program; developing a data collection plan (and assesses its feasibility); identifying the type of data collection tools required; and, addressing the logistics of data collection.

Establishing Expectations

The task in the first stage of the Step 2 is to identify the expectations for each of the evaluation questions by asking “what is the program supposed to achieve?”

To identify expectations, the following matters need to be considered for each evaluation question:

- What would satisfy you that the program has been operating successfully and achieving what was intended?
- What is the minimum that you would accept before considering making changes to the program?

When determining expectations, the following factors should be considered:

- Output versus input — What would be a reasonable output based on the input?
- Age of the program — How old is the program? Expectations may be lower for a new program than for a mature program.
- Previous experience or similar programs — How have similar programs fared? What were their results?
- Present level of outcomes — What is a reasonable change for the program to achieve?
- Realistic expectations — Set expectations which you believe are achievable.

Developing a Data Collection Plan

Selecting the Type of Tool

Creating a data collection plan is an important step in ensuring that the tools selected will provide the right data to answer the evaluation questions. The following considerations need to be addressed for each evaluation question.

- Is all the required data already available?
- What type of data collection tool would provide the data?
- Who could provide the data, if asked?
- Who can gather this data?
- What is the best design?
- From how many people or things should data be collected?
- What is the required timeframe for data collection?

The answers to these questions will inform the method for collecting the required data in a data collection plan.

Examples of data collection tools include

- Administrative records
- Registration forms
- Face-to-face Interviews
- Self-completed Questionnaires
- Telephone surveys
- Observations
- Focus Groups
- Case Studies.

Having selected the appropriate tools, the next steps are to determine who can provide the data and who can gather the data.

Design, Number, Timeframe

The evaluation design, the number of people or things data is collected from, and the timeframe for the evaluation are closely linked elements in the data collection plan. These considerations are different for process and outcome evaluation.

Process evaluation examines how a program is operating and how it might be improved. It addresses questions such as:

- Is implementation consistent with the way the program was planned?
- How can the program be improved?

The design of a process evaluation depends on whether:

- all the people or things provide data on an ongoing basis;
- a sample of people or things provide data on an ongoing basis; or
- all or a sample of the people or things provide data at specific times.

Outcome evaluation assesses the impact of a program, examining the changes that occurred as a result of a program and whether it is having its intended effect. There are many methodological issues to consider when studying questions about outcomes. Therefore, it may be appropriate to consult with a program evaluation specialist at this stage in the evaluation.

It is important to consider the various tasks involved in developing each data collection tool, then gathering and analyzing the data. A logistics plan will determine the feasibility of using the selected tools.

Step 3 – Develop Tools

Step 3 focuses on the development of data collection tools. Data collection tools are made up of measures.

Existing Measures

The first step in this stage is to determine whether existing measures will suffice. Finding existing measures and tools is challenging and time-consuming, but saves time and improves the quality of the data.

Questions, Type of Response and Response Category

If existing measures are not suitable, an existing tool will have to be modified or a new data collection tool created.

There are a variety of data collection tools – all involve asking questions and recording answers. The steps in tool development are:

- Draft the questions: Questions should be short and to the point, using simple and familiar words.
- Determine the type of response: There are two types of questions: closed-ended, which have pre-set response categories, and open-ended, which allow people to express their answers in their own words.
- Select the response categories: Select the type of response for each question; that is, recording answers using pre-set categories (closed-ended questions); or, in people's own words without pre-set response categories (open-ended questions)
- Put the questions and answers together in a clear and easy-to-understand format.

Quality Assessment

Whether new or existing tools are used, it is important to assess their quality. The evaluation is intended to produce data to help managers make decisions about the program. Managers must feel confident that their decisions are based on consistent measures. The third step at this stage involves two ways to test data quality: the content and clarity test; and the stability reliability test. The key questions are:

- Is the tool measuring exactly what it is supposed to measure?
- Is the tool easy to understand?

Step 4 – Gather and Analyze Data

Once data collection tools are developed, the next step is to gather and then analyze data. The goal at this stage is to report the facts as recorded.

Data gathering includes selecting data collectors, preparing instructions for data collection and training data collectors. It also includes pre-testing methods and assessing data collectors. This step also includes the analysis of qualitative and/or quantitative information and a summary of findings

Data Collectors and Pre-Test

The first stage in data gathering is to decide on the number of data collectors, what knowledge, traits and abilities they should possess, and how to recruit them.

A brief set of instructions for data collectors should be prepared to help maximize consistency in data collection. Depending on the type of tool and the data collector, the instructions might include:

- a description of the program that is being evaluated and from whom data will be collected;
- the purpose of the evaluation and of this particular data collection tool
- information on how to introduce and explain the tool to respondents
- instructions on how to record answers
- an outline of what they are supposed to do, when, why, where, with whom, and how
- to whom to refer respondents if the subject matter is upsetting, and
- how to answer questions that respondents are likely to ask.

It is vital that data collectors are thoroughly trained in all aspects of data collection. Keeping track of data collection activities is another important consideration.

It is important to pre-test the method and tools and assess data collectors. A pilot or pre-test is a dry run of the data collection methods.

A pre-test simulates the real thing, in advance and on a smaller scale, to detect and correct errors so there aren't any detrimental or expensive surprises later on. It is important to conduct pre-testing using the same format and data collectors, with the same type of respondents in the same setting as the actual evaluation.

In addition to pre-testing the method, it is important to assess data collectors to determine how consistently data are being collected.

Qualitative and/or Quantitative Data Analysis

Qualitative data is information collected from focus groups, interviews, and observation, plus the data collected from open-ended questions on questionnaires and surveys.

Quantitative data is information collected from the closed-ended questions on activity logs, administrative records, registration forms, interviews, surveys and records of observations.

The analysis of both qualitative and quantitative data involves a systematic, step-by-step process. Preparing for analysis involves determining the mode of analysis (i.e., by hand or computer), organizing the data, and screening for missing data and other problems.

Step 5 – Make Decisions

Step 5 requires the interpretation of the findings to inform decisions about the program and design an action plan to implement them. The process culminates with a report on the evaluation, sharing the methods and results with others.

Interpreting the Findings and Drawing Conclusions

Thinking is the most crucial element in interpreting the findings of the evaluation.

Key questions include:

- What are the patterns, what are the discrepancies?
- What internal and external factors exist?

Making Decisions and Preparing and Action Plan

The interpretation of the data will help draw conclusions about the program. These conclusions will shape the decisions. The decisions should include what changes will be made and an action plan for their implementation. Both are crucial.

Reporting on the Evaluation

It is vital that the evaluation of the program is well documented. The Tool Kit Worksheets completed during the course of the evaluation can serve this function.

It may also be important to create a written or verbal report. It should summarize the steps taken in the evaluation and take into account the audience for which it is intended.