Program evaluation utilization in Canada and its relationship to evaluation process, evaluator and decision context variables.

Michael Patrick. Shea

University of Windsor

Follow this and additional works at: https://scholar.uwindsor.ca/etd

Recommended Citation

This online database contains the full-text of PhD dissertations and Masters' theses of University of Windsor students from 1954 forward. These documents are made available for personal study and research purposes only, in accordance with the Canadian Copyright Act and the Creative Commons license—CC BY-NC-ND (Attribution, Non-Commercial, No Derivative Works). Under this license, works must always be attributed to the copyright holder (original author), cannot be used for any commercial purposes, and may not be altered. Any other use would require the permission of the copyright holder. Students may inquire about withdrawing their dissertation and/or thesis from this database. For additional inquiries, please contact the repository administrator via email (scholarship@uwindsor.ca) or by telephone at 519-253-3000ext. 3208.
NOTICE

The quality of this microform is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

If pages are missing, contact the university which granted the degree.

Some pages may have indistinct print especially if the original pages were typed with a poor typewriter ribbon or if the university sent us an inferior photocopy.

Reproduction in full or in part of this microform is governed by the Canadian Copyright Act, R.S.C. 1970, c. C-30, and subsequent amendments.

AVIS

La qualité de cette microforme dépend grandement de la qualité de la thèse soumise au microfilming. Nous avons tout fait pour assurer une qualité supérieure de reproduction.

S'il manque des pages, veuillez communiquer avec l'université qui a conféré le grade.

La qualité d'impression de certaines pages peut laisser à désirer, surtout si les pages originales ont été dactylographiées à l'aide d'un ruban usé ou si l'université nous a fait parvenir une photocopie de qualité inférieure.

La reproduction, même partielle, de cette microforme est soumise à la Loi canadienne sur le droit d'auteur, SRC 1970, c. C-30, et ses amendements subséquents.
PROGRAM EVALUATION UTILIZATION IN CANADA
AND ITS RELATIONSHIP TO
EVALUATION PROCESS, EVALUATOR AND DECISION CONTEXT VARIABLES

by

Michael P. Shea

B.A. (Hons) Simon Fraser University, 1976
M.A. Simon Fraser University, 1982

A Dissertation
Submitted to the Faculty of Graduate Studies
through the Department of Psychology
in Partial Fulfillment of the
Requirements for the Degree
of Doctor of Philosophy at
the University of Windsor
Windsor, Ontario, Canada
1991

© Copyright Michael Patrick Shea 1991.
The author has granted an irrevocable non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of his/her thesis by any means and in any form or format, making this thesis available to interested persons.

The author retains ownership of the copyright in his/her thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without his/her permission.

ABSTRACT

Data on the extent and type of program evaluation utilization in a sample of recent Canadian program evaluations were analyzed and then assessed with the characteristics of the evaluation processes employed, the evaluators who conducted the evaluations and the decision contexts in which the evaluations were performed. Surveys were mailed to all members of the Canadian Evaluation Society (CES) and the Program Evaluation Section of the Canadian Psychological Association (PES-CPA) during the fall of 1988. These individuals were followed up using procedures prescribed by Dillman (1978) resulting in a response rate of 68%. Additional data were obtained by mailing surveys to the department heads of 135 Canadian university departments. This mailing was not followed up and produced a response rate of 21%. A total of 332 completed surveys was returned from all sources. Program evaluation utilization was reported to be very extensive with 98% percent of this sample reporting utilization of some type. The most frequently occurring type of utilization was conceptual utilization (65%), followed by instrumental utilization (55%) and persuasive utilization (40%). A single type of utilization was reported in 47% of cases, two types in 42% of cases and all three in 10% of cases. The occurrence of each of the three types of utilization was related to different patterns of the evaluation process, evaluator and decision context variables.
with no one group of variables dominating the analyses. Finally, findings from research on correlates of American evaluation utilization were not replicated in the present study. There was partial support for the Patton et al. (1978) finding that the identification of a key decision maker to take responsibility for utilization was a key factor in determining utilization. Implications for evaluation practice and suggestions for future research were discussed within the context of the current study's results.
ACKNOWLEDGEMENTS

First and foremost, I would like to acknowledge the tremendous quantities of support and assistance that I received from my senior supervisor, Dr. Shelagh Towson. Secondly, I would like to acknowledge the assistance of the other members of my dissertation committee, Drs. John La Gaipa and Ray Daly of the Psychology Department and Dr. Michael Holosko of the School of Social Work. In addition, I would like to acknowledge the work of my external examiner, Dr. Mark Pancer of Wilfrid Laurier University.

I also wish to thank those members of the Canadian Evaluation Society, the Program Evaluation Section of the Canadian Psychological Association and the various Canadian university departments who gave me their valuable time by completing and returning my survey. Thanks also to the CES and CPA staff members who helped track down addresses and to the university department heads who assisted me by distributing my surveys to their department members.

I would like to thank my friends and colleagues who provided the vital support and friendship that made a return to university bearable. Special thanks also go to Ian and Nicole Knowles for their assistance in the massive mail-out. Thanks also to Dr. Robert Orr and the University of Windsor Psychology Department for some financial support of the mailing.
Thanks must also go to Dan Salhani, my former supervisor at the Ontario Ministry of Community and Social Services for all of his support and encouragement. Completing your dissertation while doing a full-time job requires all the support and encouragement you can muster. Dan was great at it.

Finally, I wish to thank my wife and partner, Alexandra Lesia Shea, for all her support and patience and for uprooting herself to move to a new community so that I could attend the program that best fit my career goals.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABSTRACT</td>
<td>iv</td>
</tr>
<tr>
<td>ACKNOWLEDGEMENTS</td>
<td>vi</td>
</tr>
<tr>
<td>TABLE OF CONTENTS</td>
<td>viii</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>xii</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>xiv</td>
</tr>
<tr>
<td>CHAPTER</td>
<td></td>
</tr>
<tr>
<td>I  INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>Three Critical Questions</td>
<td>2</td>
</tr>
<tr>
<td>Program Evaluation Utilization Type</td>
<td>3</td>
</tr>
<tr>
<td>Extent of Program Evaluation Utilization</td>
<td>5</td>
</tr>
<tr>
<td>Factors Thought to Influence Program Evaluation Utilization</td>
<td>7</td>
</tr>
<tr>
<td>The Current Study</td>
<td>10</td>
</tr>
<tr>
<td>The History of Program Evaluation</td>
<td>11</td>
</tr>
<tr>
<td>Program Evaluation Utilization Defined</td>
<td>15</td>
</tr>
<tr>
<td>The Importance of Utilization</td>
<td>20</td>
</tr>
<tr>
<td>Evidence of Program Evaluation Utilization</td>
<td>21</td>
</tr>
<tr>
<td>Factors Influencing Evaluation Utilization</td>
<td>27</td>
</tr>
</tbody>
</table>

viii
Evaluation Process Characteristics 30
Evaluator Characteristics 47
Decision Context Characteristics 53

Problem Statement 66
Research Objectives and Hypotheses 69

II METHOD 73
Subjects 73
Data Collection Procedure 73
Total Design Method 74
University Mailing Method 75
Program Evaluation Utilization Survey 76
Screening Questions 77
Dependent Variables 79
Independent Variables 81
Evaluation Process Characteristics 89
Evaluator Characteristics 91
Decision Context Characteristics 93

III RESULTS 97
Extent and Type of Program Evaluation Utilization 97
Variables Associated with Occurrence of Program Evaluation Utilization Type 105
# LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Factors Hypothesized to Influence Utilization Utilization by Category</td>
</tr>
<tr>
<td>2</td>
<td>Evaluation Process Variables Related to Program Evaluation Utilization</td>
</tr>
<tr>
<td>3</td>
<td>Evaluator Variables Related to Program Evaluation Utilization</td>
</tr>
<tr>
<td>4</td>
<td>Decision Context Variables Related to Program Evaluation Utilization</td>
</tr>
<tr>
<td>5</td>
<td>Relationships Between Evaluation Process Variables and Program Evaluation Utilization</td>
</tr>
<tr>
<td>6</td>
<td>Relationships Between Evaluator Variables and Program Evaluation Utilization</td>
</tr>
<tr>
<td>7</td>
<td>Relationships Between Decision Context Variables and Program Evaluation Utilization</td>
</tr>
<tr>
<td>8</td>
<td>Extent of Recent Canadian Program Evaluation Utilization by Type</td>
</tr>
<tr>
<td>9</td>
<td>Frequency of Different Utilization Types</td>
</tr>
<tr>
<td>10</td>
<td>Weeks Utilization Scale: Item Response Frequencies</td>
</tr>
<tr>
<td>11</td>
<td>Johnson Utilization Scale: Total Number of Items Endorsed</td>
</tr>
<tr>
<td>12</td>
<td>Chi Square Analysis of Utilization by Type</td>
</tr>
<tr>
<td>Table</td>
<td>Description</td>
</tr>
<tr>
<td>-------</td>
<td>-----------------------------------------------------------------------------</td>
</tr>
<tr>
<td>13</td>
<td>Utilization Type and Nominal-Level and Dichotomous Evaluation Process, Evaluator and Decision Context Variables</td>
</tr>
<tr>
<td>14</td>
<td>Occurrence of Utilization and Interval-Level Evaluation Process, Evaluator and Decision Context Variables</td>
</tr>
<tr>
<td>15</td>
<td>Variables Used in the Three MANOVAS</td>
</tr>
<tr>
<td>16</td>
<td>Utilization and Identification of Key Decision Makers and Evaluation Users</td>
</tr>
<tr>
<td>G-1</td>
<td>Total Design Method Response Rates by Response Category</td>
</tr>
<tr>
<td>G-2</td>
<td>Response Frequencies by Type of University Department</td>
</tr>
<tr>
<td>G-3</td>
<td>Response Categories by Survey Method</td>
</tr>
<tr>
<td>H-1</td>
<td>Comparison of CES and PES-CPA Respondents and Nonrespondents</td>
</tr>
</tbody>
</table>
**LIST OF FIGURES**

<table>
<thead>
<tr>
<th>Figure</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>116</td>
</tr>
<tr>
<td>Patterns of Association Between Utilization Type and Evaluation Process, Evaluator and Decision Context Variables</td>
<td></td>
</tr>
</tbody>
</table>
CHAPTER I
INTRODUCTION

Program evaluation is a type of social scientific inquiry whose primary goal is to provide information on the efficiency and effectiveness of social, health, education and other types of human service programs (Cook, Leviton, & Shadish, 1985; Patton, 1986). Program evaluation does not rely on any one social scientific method; it draws freely from all types including both quantitative and qualitative methods as needed. In most cases, the information obtained from program evaluation research findings is intended to be used by program decision makers to improve the performance of their programs. Therefore, according to many program evaluators (Alkin & Coyle, 1988; Cook & Shadish, 1986; Patton, 1987), program evaluations which produce findings that are not utilized by their intended audiences cannot be considered to be successful.

Until the mid 1960s, program evaluation was not generally viewed as a distinct discipline and program evaluations were relatively rare events with the exception of evaluations of some educational and military programs. However, since then, a number of factors have combined to help define, shape and bolster the discipline of program evaluation, particularly within government-funded programs. In fact, a program evaluation is now required in many
instances in both Canada and the United States in order for a program to maintain its funding (Rutman & Mayne, 1985; Wholey, 1986).

During the 1970s and early 1980s, program evaluation became increasingly defined as a distinct discipline of social scientific inquiry (Cook, 1978; Cook & Shadish, 1986; Hudson, 1988). In both Canada and the United States, program evaluators formed professional associations, began publishing professional journals dedicated solely to program evaluation issues and studies and developed standards for evaluation practice (Rossi, 1982).

The prospects for evaluation and evaluators have been positive in recent years. Increasingly, university programs of graduate study in professional disciplines such as psychology, educational administration and social work have been providing program evaluation training for their students. Furthermore, a number of graduate programs have been founded specifically to train program evaluation researchers.

Three Critical Questions

Since program evaluation first emerged as a distinct area of social scientific inquiry in the mid 1960s, three questions have been asked repeatedly by program evaluation
theorists, practitioners and consumers (Agarawala-Rogers, 1977; Alkin & Coyle, 1988; Braskamp & Brown, 1980; Caro, 1971; Ciarlo, 1981; Leviton & Hughes, 1981; Patton, 1987; Rich, 1977; Suchman, 1967; Weiss, 1966; Weiss, 1972). These questions are: What types of program evaluation utilization are occurring? How much program evaluation utilization is actually occurring? Which factors influence or determine those instances of program evaluation utilization that do occur? As the discipline of program evaluation enters its third decade of heightened activity, the answers obtained to these three questions have a seemingly important bearing on its future (Cook, 1984; Hudson, 1988; Patton, 1987).

Program Evaluation Utilization Type

It is important that program evaluators agree on what constitutes program evaluation utilization. Empirical research and theory on this issue has been hampered by the continuing definitional debates. For instance, as Beyor and Trice (1982) and later Cousins and Leithwood (1986) observed, after reviewing the empirical research literature on program evaluation utilization, the lack of common operational definitions has seriously reduced the comparability of the results of different research studies of utilization.
Suchman (1967), Weiss (1972), Rich (1977) and Leviton and Hughes (1981) were apparently at the forefront of the debate over how to define program evaluation utilization. Initially, these pioneers perceived evaluation utilization solely as the direct, instrumental use of evaluation findings by program decision makers. Thus, the basic decision to be faced at the end of the evaluation was whether or not to continue the program. Later, Weiss (1972) argued that program evaluation utilization of this type was relatively rare and that indirect, conceptual uses were more common. Cumulative (evaluation) experience has demonstrated that very few programs are terminated solely as a result of their poor performance being demonstrated by an evaluation study (Patton, 1986).

Rich (1977) studied utilization of the results of social scientific research by federal bureaucrats in the United States and found that utilization of two types had occurred. He confirmed that research results are sometimes used in a direct, instrumental manner to make decisions that affect program operations. In addition, he observed that many research results are used in the indirect, conceptual manner described by Weiss (1972) and others. Similarly, Leviton and Hughes (1981) reviewed the literature on program evaluation utilization and concluded that the existence of both instrumental and conceptual utilization had been demonstrated empirically. They also
speculated that a third type of program evaluation utilization existed, persuasive utilization, in which program evaluation results are utilized to defend or attack the program which underwent the evaluation. However, they noted that the existence of this third type of program evaluation utilization had not been demonstrated empirically. The debate over which types of program evaluation utilization exist continues at present (Cousins & Leithwood, 1986). No research has yet been conducted which has demonstrated the distinction among these three different types of program evaluation utilization.

Extent of Program Evaluation Utilization

Speculation with regard to the extent of program evaluation utilization began in the mid 1960s and has continued to the present (e.g., Dunn, 1983; Leviton & Hughes, 1981; Shea, 1988a; Weiss, 1966). However, few empirical studies have been conducted on this subject. This concern is of particular importance to the discipline of program evaluation because as Cook (1984), Patton (1987) and others have observed, the perception that program evaluations are not producing utilizable results may cause funding bodies to reduce the resources available for the conduct of program evaluation research.
As early as 1966, Weiss was questioning the extent to which the results of program evaluations were actually utilized by program decision makers to improve program performance. She noted that, in her experience, instances of the direct use of evaluation findings in making major decisions about programs were indeed very rare (Weiss, 1966).

These doubts about utilization remained largely quiescent during the latter part of the 1960s and early part of the 1970s, although some concern was expressed by Schulberg and Baker (1971) and Bigelow and Ciarlo (1976). It was left to researchers such as Patton and his colleagues (Patton, Grimes, Guthrie, Brennan, French & Blyth, 1978) to revitalize the debate and bring empirical data to bear on this question. Later, Patton et al. (1978) re-confirmed the notion that instances of direct, instrumental utilization of evaluation findings were relatively rare. However, they also found that indirect, conceptual uses of evaluation findings were quite common.

Speculation regarding program evaluation utilization has continued unabated in the 1980s. Unfortunately, this debate has been primarily theoretical, as empirical investigations of the extent of utilization have been relatively scarce (Cousins & Leithwood, 1986; Leviton & Hughes, 1981). In addition, there are no systematic data on the extent of evaluation utilization in Canada.
Factors Thought to Influence Program Evaluation Utilization

Speculation regarding which factors influence or determine program evaluation utilization in the evaluation literature began in the 1960s and continues presently (e.g., Davis & Salasin, 1975; Rossi, 1966; Rog & Bickman, 1984). By contrast, empirical investigations of this topic did not commence until the mid 1970s and are far less common (e.g., Johnson, 1980; Patton et al., 1978; Weeks, 1979).

In the realm of this speculation, Rossi (1966) hypothesized that the degree of experimental rigour in the evaluation's research design was the key determinant of program evaluation utilization. Conversely, Davis and Salasin (1975), and Rog and Bickman (1984) have argued that particular styles of program evaluation research lead to greater evaluation utilization. These authors have described quite different styles of evaluation processes designed to enhance such utilization.

In the empirical realm, Patton et al. (1978) observed that utilization occurred to a greater extent in evaluations in which the evaluator had identified a specific program decision maker who would take personal responsibility for ensuring that utilization occurred. Further, Weeks (1979) found that evaluations which employed multiple measures of program performance had a higher
probability of being utilized. Finally, Johnson (1980) reported that the presence of information brokers within the decision context in which the evaluation was conducted was related to greater utilization.

The belief that evaluation results are underutilized has led some authors such as Dunn (1983) to conclude that evaluation research is obsolete and that resources for its conduct should be reduced. As Cook (1984) noted, it is vital that program evaluation consumers be disabused of this belief if the future of the discipline is to be assured.

The importance of the utilization issue for evaluators was underscored in a recent interview with Michael Patton, then President of the American Evaluation Association (Hudson, 1988) who stated, "We (evaluators) are evaluated in terms of our utility, our practicality, and our accuracy ... what we do ought to make a difference, it ought to be used and if it is not, no matter how elegant the methods, no matter how rigorously we have done the work, if evaluations are not getting used, we are not doing the job" (p. 123).

Obtaining answers to the three evaluation utilization questions previously posed seems important because utilization success is the single most important criterion for judging program evaluation research. Conversely, an evaluation study which produces results that are not
utilized by its intended audiences is deemed a failure and a waste of resources. It is, therefore, crucial that evaluators know how much and what types of program evaluation utilization are occurring and communicate objective information about this issue to potential funders of evaluation research activities.

In addition, many program evaluators believe that it is necessary to understand which factors influence or determine utilization so that their evaluation practice can be tailored accordingly, thus increasing the probability that their evaluation findings will be utilized (Brown & Braskamp, 1980; Cook, Leviton, & Shadish, 1985; Pancer, 1985; Patton, 1986; Shea, 1988b).

At present, the three program evaluation utilization questions remain unanswered to a large extent, especially in Canada. One result of this is that program evaluators have yet to agree on which evaluation practice approaches are more likely to lead to utilization or even on what constitutes utilization per se. Most of the empirical studies which exist are based on a small number of case studies or simulated evaluations (Beyer & Trice, 1982; Cousins & Leithwood, 1986). Fortunately, a small but growing body of empirical research studies on program evaluation utilization in the United States has demonstrated that utilization is far more common than was previously believed. In addition, various researchers
(e.g., Alkin et al., 1979; Johnson, 1980; Patton et al., 1978; Weeks, 1979) have begun to delineate some of the factors which determine utilization success. These studies will be discussed later in this investigation.

Patton's (1988) statement regarding the importance of a concern for program evaluation utilization, published in a Canadian evaluation journal, seems as true for Canadian as for American program evaluators. At the present time however, Canadian program evaluators have little systematic information regarding either the extent to which or the ways in which their evaluations are being utilized. This lack of concrete information is especially important given the fact that a number of Canadian program evaluators have hypothesized that the Canadian evaluation context differs in some key ways from that in the United States, thus reducing the potential applicability of American research findings in Canada (Rutman, 1986; Rutman & Mayne, 1985).

The Current Study

The research reported in this study represents a first step in the provision of information on the extent and types of program evaluation utilization success in Canada and on factors related to instances of utilization success. In addition, the current study represents the first attempt
to demonstrate the distinctiveness of the three different types of program evaluation utilization modes which have been described in the literature. Data on these questions were collected by means of a survey of a sample of Canadian program evaluators. These data included measures of program evaluation utilization and a number of evaluator, evaluation, and decision context variables hypothesized to determine utilization.

Before proceeding to a more detailed discussion of the objectives and hypotheses of the current study as well as its methodology and results, a review of the literature on program evaluation and program evaluation utilization will be presented.

The History of Program Evaluation

Although individual evaluation research studies had been conducted earlier, the development of program evaluation research began in earnest with the rapid growth of federally-funded government social services in the United States during the depression of the 1930s. Through the New Deal initiatives, the federal government of the United States took unprecedented responsibility for ensuring the social welfare of a large number of its citizens. The resulting massive increases in government
expenditures for social programs were accompanied by a call for greater public accountability. Rossi and Freeman (1982) credited A.S. Stephan (1935) as being among the first to call for the application of experimental methods to the study of the effectiveness and efficiency of government social programs. Stephan and other American social scientists of the pre World War II era argued that the effectiveness of specific social interventions should be demonstrated on a small scale before they were applied nationally. They further argued that evaluations of the effectiveness and cost efficiency of social programs should be conducted using experimental approaches in order to ensure the validity of conclusions reached.

At the end of World War II, American government social welfare services again expanded greatly, this time to meet the needs of returning war veterans, resulting in the creation of further opportunities for evaluation research. Especially within the education field, evaluation studies became commonplace, promoted primarily by the work of Tyler (1950) and others.

During this early evolution of program evaluation research, financial support was inconsistent. Funds for evaluation research were usually tied to specific programs as opposed to the current situation in which "generic" funds for evaluation are often made directly available to the evaluation units of government departments. At this
time, systematic and scientific program evaluation remained the exception rather than the rule. However, by the 1960s program evaluation began to emerge as a legitimate area of social scientific study, and evaluation studies became more common. The publication of Campbell and Stanley’s (1966) classic work on the strengths and weaknesses of various experimental and quasi-experimental research designs and Suchman’s (1967) review of evaluation research methods and issues were milestones in the early development of the area. Subsequent key events included the publication of Struening and Guttentag’s (1975) Handbook of Evaluation Research and the first volume of the Evaluation Studies Review Annual (Glass, 1977). The founding, in the mid 1970s, of the Evaluation Network, the Evaluation Research Society and the Evaluation Review (1977) signalled the legitimization of program evaluation as a distinct area of social scientific activity. Further evidence for the developing maturity of the area was the creation of standards for evaluation practice (Rossi, 1982) and the inclusion of a chapter on program evaluation by Cook, Leviton, and Shadish in the 1985 edition of The Handbook of Social Psychology. The amalgamation of the Evaluation Research Society and the Evaluation Network into the American Evaluation Association in the late 1980s represented the culmination of these trends.
With this legitimization process also came increasing financial support for evaluation research and the creation of relatively stable financial commitments for the funding of evaluation studies. For example, the American federal government's Community Mental Health Act of 1964 required that all federally-funded Community Mental Health Centers spend a minimum of two percent of total program budgets on evaluation activities. In addition, Thompson (1975) argued that President Johnson's establishment of the Planning, Programming and Budgeting System (PPBS) for U.S. federal government programs in 1965 was a key event in establishing a solid base of financial support for evaluation research in the United States.

Within the last decade, the Canadian federal government has mandated and funded a broad based system of program evaluation activity. Evaluation activities are monitored and evaluated through the Office of the Comptroller General. Rutman and Mayne (1985) reported that as of March 31, 1984, 42 Canadian federal government departments had permanent evaluation units. These 42 departmentally-based evaluation units, housed in government departments with responsibility for over 85% of federal government expenditures, employed over three hundred full-time program evaluators and spent an additional 10.9 million dollars on non-salary expenditures. The number of program evaluators employed by the Canadian federal
government and the funds allocated to program evaluations of federal programs have continued to increase since that time.

In addition, program evaluators in Canada have developed a professional association. The Canadian Evaluation Society hosts national and provincial conferences, publishes its own journal and is currently developing its own code of professional conduct.

Program Evaluation Utilization Defined

The definition of program evaluation utilization has changed considerably over time as has the definition of program evaluation as a whole. During the early days (1960s), program evaluation was defined primarily as another application of social science methodology, and its focus was almost exclusively on the assessment of program outcomes. For example, Suchman (1967) defined evaluative research as the utilization of scientific research methods for the purposes of making an evaluation or judgment of worth with regard to a social program. Not surprisingly, Suchman’s perception of utilization was also relatively narrow, the direct use of evaluation findings by program administrators in determining actual decisions about a particular program’s operations. Indeed, this definition
describes a direct, instrumental approach to this issue.

Subsequently, Weiss's (1972) definition of both evaluation research and evaluation utilization was broader, as she defined evaluation research as a systematic process for rendering judgments about the success of social programs designed to improve quality of life. In addition to noting the utility of the tools of social science research in program evaluation research, Weiss also noted that the context in which evaluation studies are conducted places unique and difficult demands on social science methodology which require the specific attention of evaluators.

Struening and Guttentag's (1975) work builds on Weiss's (1972). More specifically, they outlined ten crucial program evaluation steps: conceptualizing the problem, reviewing relevant literature, developing a research strategy, determining a research design, selecting and maintaining a sample, choosing measures and assessing their psychometric properties, selecting appropriate personnel to conduct the study, maintaining data collection standards, analyzing the data, and communicating the results.

With reference to evaluation utilization, Weiss defined it as occurring only when the results of an evaluation contributed to subsequent program decision making and the improvement of future programming. According
to Weiss, these potential contributions denoted the social purpose of program evaluation research. Indeed, this view recognized the indirect, conceptual nature of some aspects of program evaluation utilization but was still somewhat restrictive.

By the late 1970s and 1980s, program evaluation was regarded as a distinct area of social scientific inquiry, and the definition of evaluation utilization had expanded to include both instrumental and conceptual uses. The evaluation of the conception of program evaluation as a distinct area is reflected in the work of Rossi and Freeman (1985), Posavac and Carey (1985) and Cook and Shadish (1986).

Rossi and Freeman (1985) defined evaluation research as the systematic application of social research procedures in assessing the conceptualization and design, implementation, and utility of social intervention programs. In their view, program evaluation involves more than the application of social scientific methods. It is also seen as a political and managerial activity and as one of many inputs into the complex process of program decision making. Similarly, Posavac and Carey (1985) stated that the mission of program evaluation is to assist in improving the quality of human services and defined evaluation as a collection of methods, skills, and sensitivities necessary to determine whether a human service is needed and likely
to be used, sufficiently powerful to meet the identified need, offered as planned, and capable of actually helping people in need.

Finally, Cook and Shadish (1986), in a comprehensive review of the field, labelled program evaluation as the worldly science. In their view, evaluations are conducted in order to produce knowledge about the value of social programs and their constituent parts, knowledge that can be used to make programs more responsive to the social problems they are designed to ameliorate. They contended that a comprehensive theory of program evaluation requires a knowledge base about social programs, evaluation usage, the assignment of value and constructing valid knowledge. They argue further that these four knowledge bases must converge into a theory of evaluation practice that guides question and method choice within the limits set by budget, time, and staff constraints, and in the light of trade-offs that inevitably follow once particular method or issue choices have been made.

With regard to the elaboration of the definition of program evaluation utilization, the focus of the present investigation, Rich (1977) was the first to expand on the work of Suchman (1967) and Weiss (1972). According to Rich, instrumental use occurred when research findings were cited specifically in making a program or policy decision. Conceptual use occurs when research findings influence the
thinking of policy makers who then make plans for future use of the information. Rich contended that both types of use were important and that the conceptual use of evaluation findings was especially important during the process of program planning and development. This author was the first to delineate both instrumental and conceptual program evaluation utilization empirically and to emphasize the importance of the latter approach. Subsequent research suggests that the issue is even more complex than Rich (1977) had imagined.

Based on their review of the theoretical and empirical literature on program evaluation utilization, Leviton and Hughes (1981) observed that Rich's (1977) distinction between instrumental and conceptual utilization had been supported by a number of independent empirical investigations of actual utilization. In addition, they proposed that evaluation findings could be utilized as persuasive utilization, defined as attempts to persuade others to continue to support a program or line of action or to abandon it based upon evaluation findings. This contention was not based on empirical findings, but rather on the anecdotal reports of a number of evaluation practitioners.

Cousins and Leithwood (1986), in the most recent review of the empirical literature on program evaluation utilization, suggested that utilization had been
operationally defined in four different ways: direct support for discrete programs (instrumental utilization); education of decision makers (conceptual utilization); the mere psychological processing of evaluation results (whether or not decision makers were aware of and had thought about the results); and potential utility. However, they concluded that the instrumental and conceptual definitions of program evaluation utilization remained the most frequently used and most important with the other definitions of utilization being considerably less important. These authors did not find evidence for the existence of the persuasive utilization type postulated earlier by Leviton and Hughes (1981).

The Importance of Utilization

Increased financial support for evaluation research in both the U.S. and Canada has been accompanied by increased calls for evidence of program evaluation utilization (e.g., Caro, 1971; Conner, Altman, & Jackson, 1984; Cook, 1984; Davis & Salasin, 1975; Dunn, Mitroff, & Deutsch, 1981; Patton, 1978; Saxe & Koretz, 1982; Weiss, 1972). Evaluation theorists and practitioners have attempted to reach some agreement on how to measure the extent of program evaluation utilization success (e.g., Alkin et al., 1979;
Bigelow, 1975; Patton et al., 1978; Stevenson, 1981; Weeks, 1979). In addition, minimal agreement exists regarding the best ways to explore possible determinants of evaluation utilization (e.g., Braskamp et al., 1982; Cousins & Leithwood, 1986; Greene, 1988; Patton et al., 1978; Weeks, 1979; Weiss & Bucuvalas, 1981). A number of evaluators, program decision makers and policy makers have expressed concern that evaluations are not utilized as often as they ought to be (e.g., Agarawala-Rogers, 1977; Conner, 1981; Cook, 1984; Patton, 1978; Robinson, 1988; Schulberg & Baker, 1971; Weiss, 1966). Robinson (1988), for example, has called the program evaluation utilization issue the "Achilles heel" of program evaluation. In addition, the topic of program evaluation utilization was the central theme of the 1988 conference of the American Evaluation Association. The literature on the extent and types of program evaluation utilization success and on factors related to evaluation utilization success will be reviewed at this point.

Evidence of Program Evaluation Utilization

An overview of the literature reveals, first, that empirical investigations of program evaluation utilization are far less common than non-empirical speculations
(Leviton & Hughes, 1981). Second, of those empirical investigations which do exist, the majority suffer from methodological and definitional problems that reduce their reliability, validity, comparability, and the generalizability of their findings (Alkin, 1980; Conner, 1981; Cousins & Leithwood, 1986; Dunn, 1983; Leviton & Hughes, 1981; Weiss, 1981).

The most frequent type of utilization research consists of case studies of one or more evaluations. Of those studies which employ rigorous research designs, most are simulated situations that do not use actual evaluators or decision makers as subjects. Past studies of program evaluation utilization have also been criticized for using, as dependent variables, factors which might influence utilization rather than direct indices of evaluation utilization. Furthermore, many studies of evaluation utilization have been criticized for using measures of unknown reliability and validity (e.g., Conner, 1981; Cousins & Leithwood, 1986; Dunn, 1983; Dunn, Dukes, & Cahill, 1984; Weiss, 1981). In addition, although both Weeks (1979) and Johnson (1980) have developed evaluation utilization scales, these scales have been used in only a small number of studies. Finally, in this regard, Cousins and Leithwood (1986) have criticized past studies for considerable methodological variability in operationalization of the utilization construct.
Despite these limitations, it seems worthwhile to examine those studies which have addressed the first two research questions: what types of evaluation utilization are occurring, and to what extent are the various types occurring?

Convincing evidence exists for both instrumental and conceptual program evaluation utilization (e.g., Alkin et al., 1979; Anderson, Ciarlo, & Brodie, 1981; Bigelow, 1975; Cook, Levinson-Rose, & Pollard, 1981; Florio, Behrman, & Goltz, 1979; Greene, 1988; Johnson, 1980; Patton et al., 1978; Rayner, 1986; Rich, 1977; Siegal & Tuckel, 1985; Smith, 1976; Weeks, 1979).

Bigelow (1975) was unable to identify a clear cut instance of instrumental utilization in a case study of an American community mental health center (CMHC) but did find some evidence of conceptual utilization. Evidently, the evaluation research results were only one of the information sources used by program personnel in making decisions about the center, and thus other information was more persuasive. However, the evaluation results did provide information that was of indirect use in general program development discussions.

Bigelow's results are consistent with those of Smith (1976), who also observed that findings from a CMHC evaluation were utilized conceptually in making program development decisions and in the planning and development
of new programming alternatives.

Rich (1977) reported that federal government officials in the U.S. made many instrumental uses of social science research, with 94% of the instances of instrumental utilization occurring in the first three months following the release of the information. Instances of conceptual utilization of social scientific research findings were also common. Furthermore, conceptual utilization increased as more time elapsed following the presentation of the evaluation report.

Patton et al. (1978) collected data from evaluators and key decision makers involved in twenty evaluations of U.S. federally-funded health programs and reported that 14 of the 18 responding decision makers and 13 of the 14 responding evaluators believed that the evaluation had made an impact on the program. However, when the respondents were asked to define utilization in personally meaningful terms, not one described a clear instance of instrumental evaluation utilization. Instead, respondents described the use of evaluation results in indirect, conceptual ways, including the provision of additional information which served to reduce uncertainty in decision making processes.

Other investigators have reported that conceptual utilization is common whereas instrumental utilization is not. For example, Alkin et al. (1979) reported that only one of five evaluations of educational programs was
utilized directly in making major decisions about the evaluated program, whereas four of these same five evaluations were utilized conceptually. In addition, Florio et al. (1979) observed that only a minority of educational program evaluations actually resulted in significant program changes (instrumental utilization). These researchers questioned the extent to which education program administrators relied upon evaluation research findings in their program development decisions. Furthermore, Weeks (1979), in an extensive study of evaluation utilization, observed that only 29% of a sample of 57 program evaluations were utilized instrumentally. He also found that 49% of this same sample were utilized conceptually. According to Johnson (1980), only 11% of a sample of 75 program decision makers reported that the evaluation findings were used directly in making program modification decisions. He also reported that 89% of this same sample could be described as conceptual utilization. Siegel and Tuckel (1985) reported that conceptual program evaluation utilization occurred in two case studies of evaluation utilization within the U.S. Government General Accounting Office and in Ralph Nader's Center for the Study of Responsive Law. In a subsequent examination of utilization in the community mental health programming field, Anderson, Ciarlo and Brodie (1981) reported that program decision makers made use of approximately 45% of
the feedback statements obtained from an evaluation in one mental health program and 36% in the other. In addition, Greene (1988) observed that conceptual program evaluation utilization occurred in two recent evaluations of American social programs.

Further, Cook, Levinson-Rose and Pollard (1981) suggested that the results of some program evaluations are misutilized. These authors cited a number of instances in which program evaluation findings were used in unintended ways, for example, to attack a program that had been evaluated in an attempt to terminate its funding. These cases could be considered to be examples of misutilization or, alternatively, could be perceived as instances of persuasive program evaluation utilization (suggested by Leviton & Hughes, 1981).

In one of the few reports of utilization in a Canadian context, Rayner (1982) noted that during his years within the evaluation field of the Canadian Federal government, he had observed the utilization of over 150 evaluations of government programs. However, he did not supply detailed data on his findings.

In sum, these findings demonstrate that program evaluation findings are often utilized. Specifically, in five of the ten studies which reported data on the extent of utilization, more than 50% of the evaluations were utilized. In addition, most studies demonstrated that
program evaluation findings are utilized in a variety of ways and that conceptual utilization is by far the most common type of utilization. However, these findings also provide some indication that evaluation utilization is not inevitable. It should also be noted that these findings should be regarded as tentative given that most of them are based on a small number of case studies.

Factors Influencing Evaluation Utilization

Much of the recent literature on program evaluation utilization in the United States has attempted to delineate factors which influence evaluation utilization. Unfortunately, most of this literature has been non-empirical, with evaluators and theorists drawing upon their own experiences in conducting evaluations and also on anecdotal information. Speculation with regard to this topic, which began in the 1960s and continues at present, has been wide ranging. During program evaluation's relatively short history, different program evaluation theorists have suggested that individual factors or groups of different factors play a prominent role in determining the degree to which results of program evaluation research are utilized.
The authors of two comprehensive literature reviews have attempted to group these factors into varying numbers of categories. Leviton and Hughes (1981) proposed that five clusters of factors (or variables) influence program evaluation utilization. They contended that all of these factors working together explain the great variability in evaluation utilization. Three of these clusters of variables are evaluation process characteristics, one is a characteristic of the evaluator, and one is a characteristic of the decision context in which the evaluation is conducted. The evaluation process characteristics are: the relevance of the findings to the needs of potential users; the extent of communication between the evaluator and potential users of the evaluation findings; and, the translation of evaluation findings into recommendations for program or policy change. The evaluator characteristic is the amount of trust placed in the evaluator by potential users (as distinguished from the amount of credibility placed in the evaluation process). Finally, the decision context characteristic is the degree of commitment or advocacy for the use of the evaluation's findings evidenced by individual users who are program decision makers.

In a subsequent review of only the empirical literature, Cousins and Leithwood (1986) concluded that 12 clusters of variables similarly influenced utilization.
These were: evaluation quality, credibility, relevance, communication quality, valence of findings, timeliness, information needs, decision characteristics, political climate, competing information, personal characteristics of decision makers, level of commitment and/or receptiveness to the evaluation of program personnel. Cousins and Leithwood (1986) dichotomized these variables into two groups which they labelled characteristics of the evaluation implementation and characteristics of the decision or policy setting in which the evaluation was conducted. Not addressed in the scheme were such potentially influential evaluator characteristics as the evaluator's title (Braskamp, Brown & Newman, 1978), the evaluator's gender (Newman, Brown & Litman, 1979) and the evaluator's personal credibility (Weiss & Bucuvalas, 1980).

This dichotomous grouping by Cousins and Leithwood (1986) is at odds with the position being taken in the current study. On the basis of the existing evidence, it is the perception of the author that the factors believed to influence the extent and type of program evaluation utilization can be viewed as belonging to three major categories: characteristics of the program evaluation process; characteristics of the program evaluator; and, characteristics of the decision context in which the program evaluation was performed. The assumption advanced here is that each of these three categories includes
distinct and mutually exclusive clusters of variables which may play important roles in determining the eventual level of utilization of evaluation findings. The following discussion focuses on non-empirical speculations and empirical findings with regard to each of the factors believed important in these three categories. These factors are listed by category in Table 1.

**Evaluation Process Characteristics**

Most of the speculative literature on evaluation utilization determinants has indicated that evaluation process characteristics play the dominant role in determining program evaluation utilization. Specifically, such characteristics are defined as features or components related to the way in which a program evaluation study is conducted. In this regard, process characteristics which have been suggested by various authors include the level of methodological rigor of the evaluation research design, the credibility of the evaluation process, the style of program evaluation practice employed by the evaluator and the timeliness of the evaluation.

Rafter (1984) noted that there have been three main approaches to program evaluation research: the scientific approach which stresses the importance of the production of
Table 1

Factors Hypothesized to Influence Utilization by Category

<table>
<thead>
<tr>
<th>Evaluation Process</th>
<th>Evaluator</th>
<th>Decision Context</th>
</tr>
</thead>
<tbody>
<tr>
<td>Methodological Rigor*</td>
<td>Understanding of the Program</td>
<td>Decision Making Style</td>
</tr>
<tr>
<td>Credibility</td>
<td>Communication Ability*</td>
<td>Perception of Threat by Staff</td>
</tr>
<tr>
<td>Evaluation Practice Style*</td>
<td>Organizational Position*</td>
<td>Resistance to Change</td>
</tr>
<tr>
<td>Timeliness of Results*</td>
<td>Personal Credibility</td>
<td>Decision Makers’ Reason for Initiation*</td>
</tr>
<tr>
<td></td>
<td>Level of Training*</td>
<td></td>
</tr>
<tr>
<td>Evaluation Focus*</td>
<td></td>
<td>Blocked Feedback</td>
</tr>
<tr>
<td>Direction of Evaluation Results*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lack of Specific Recommendations in Report*</td>
<td></td>
<td>Competing Information*</td>
</tr>
<tr>
<td>Insufficient Evaluator Time on Dissemination*</td>
<td></td>
<td>Decision Makers’ Organizational Position</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Lack of Clear, Measurable Objectives</td>
</tr>
</tbody>
</table>

Note: (*) These variables are included in the current study.
high quality methodology; the interactive approach which stresses the importance of interactions with program stakeholders; and, the hybrid approach which recognizes the importance of both issues for ensuring utilization. He hypothesized that evaluators who practised the hybrid approach would have a greater likelihood of having their evaluation findings utilized than those evaluators who practised the other approaches.

Rossi (1966) and Suchman (1967) clearly subscribe to the first of Rafters' approaches. Both researchers contended that the probability of program evaluation utilization could be increased by increasing the methodological rigor of evaluation research designs. A true experimental research design (Campbell & Stanley, 1966), in which subjects are assigned randomly to experimental and control groups, was generally considered to be the ideal scenario in this regard. From this perspective, Berk and Rossi (1977) reported that the scientific authority of a well formulated evaluation research design produced the greatest impact on evaluation utilization success rates. Further, they believed that this scientific authority gave an evaluation overall credibility in the eyes of program decision makers and policy makers and that this credibility was the key determinant of an evaluation's eventual utilization.
Many program evaluators seem to subscribe to the interactive approach described by Rafter (1984), arguing that the extent of program evaluation utilization may be increased by practising a particular evaluation style (Cronbach, 1982; Davis & Salasin, 1975; Heilman, 1983; Neigher, 1979; Pancer, 1985; Patton, 1978; Rafter, 1984; Rich, 1979; Rog & Bickman, 1984; Rutman, 1980; Rutman; 1986). For example, Davis and Salasin (1975) contended that an evaluator who adopted the role of change agent as opposed to the role of value-neutral, methodological expert would be more likely to attain utilization. They described in detail an evaluation approach which they believed would enhance the likelihood of achieving program evaluation utilization. This approach involves the clear specification of program goals and objectives and the involvement of program decision makers in the utilization process.

Similarly, Patton’s (1978) Utilization-Focused Evaluation Model prescribes involvement of program stakeholders in all phases of the evaluation process including planning, implementation and utilization. Patton noted that the single best method to enhance the utilization of the results of any evaluation is to ensure that intended users and intended uses of the information produced by an evaluation are identified prior to its commencement. He also contended that if it is not possible to identify intended users and intended uses then the
evaluation should not be performed.

Neigher (1979) observed that certain evaluation activities have been more successful than others in providing useful information to the staff of U.S. mental health programs. Specifically, he noted that whereas needs assessments and management information systems (MISs) are likely to produce useful information, outcome evaluations are less likely to be utilized. He recommended that evaluators concentrate their efforts on the evaluation of program processes and that program staff be consulted prior to the design of information systems. Similarly, although Rich (1979) believed that it may be unwise to attempt to identify a single best method for evaluation in the hopes of enhancing evaluation utilization, he agreed that program managers should play an active role in evaluation planning and design. Pancer (1985) also recommended that evaluators use a participative approach to evaluation research, if they hope to have their evaluation findings utilized by program stakeholders.

Like Patton (1978), Rutman (1980) observed that not all programs are equally ready to be evaluated. He suggested that an evaluability assessment should be performed prior to the decision to evaluate a program and provided methodology designed to facilitate this task. Rutman indicated that programs judged as being more evaluable by this process would be more likely to produce
evaluation results that were utilized.

Rutman (1986) cited evidence that many evaluations of Canadian federal government programs were useless, and he contended that the main reason for utilization failure in these cases was the failure of the program evaluation to examine the underlying rationale of the program. He argued that maximum utilizability would be produced by evaluations which have a clear and specific focus and produce relevant results on an ongoing basis. Rutman stated that despite the belief of many evaluators, the methodological rigor of the evaluation research design is not a primary determinant of utilization. Similarly, Heilman (1983) predicted that evaluators who produced synthesizing reviews of past evaluation findings would probably obtain high levels of utilization of their findings, since evidence produced by a series of evaluation studies would have greater weight than that produced by a single study.

According to Levitan and Wurzberg (1979), the central issue in explaining the relatively low rates of program evaluation utilization is the manner in which most evaluations are performed. They argued that most evaluations employ research designs of such low rigor that evaluation as practised fails to provide a convincing or credible basis for making program decisions. Further, they recommended that evaluators alter their methods to more appropriately meet the needs of program decision makers.
Based upon extensive evaluation experiences, Cronbach (1982) observed that ensuring that the evaluation is of the highest methodological quality does not necessarily guarantee that an evaluation will move its intended audiences to action. He contended that other aspects of the process of conducting program evaluations were equally important in determining eventual utilization. For example, he indicated that evaluations which examined identifiable and transportable elements of programs are more likely to be utilized than those which examined overall program performance. In addition, program evaluations which focus on program processes rather than strictly on program outcomes are more likely to be utilized.

Rog and Bickman (1984) also noted that evaluators who have let methodological rigor serve as their main guide to enhancing utilization have often been disappointed. They recommended that evaluators practice an approach to evaluation that relies heavily on feedback from program stakeholders at all stages of the evaluation process. They further indicated that the use of this approach would enhance the likelihood of utilization of the evaluation's results by its intended audiences.

Guttentag (1977) and Rayner (1986) made claims regarding the importance of timing in determining the extent of program evaluation utilization. Guttentag contended that information provided by evaluators to
program decision makers must be timely in order to be fully utilized as intended. In her view, timely evaluation results are those which are produced in time for program decision makers to use them to meet their decision deadlines. Similarly, Rayner (1986) indicated that the key determinant of the extent of program evaluation utilization within the Canadian federal government was the timeliness with which evaluation results were made available in relation to the program decision making process. He defined timely evaluation results as those that are produced on an ongoing and predictable basis.

Rather than identifying a single evaluation process characteristic as the primary determinant of the extent of evaluation utilization, a number of authors have speculated that a group of evaluation process characteristics working together determine utilization extent. Weiss (1966) suggested that the direction of the evaluation's results and the manner in which the evaluation was performed might influence evaluation utilization and, as a result, questioned Rossi's (1966) assertion that the methodological quality of the evaluation research design would be the primary determinant of utilization. Agarawala-Rogers (1977) observed that program evaluations are seldom utilized as intended in the improvement of program operations. She believed that this was the result of many factors which included: a lack of administrative participation in the
planning of the evaluation; the conflicting interests of program evaluators and program decision makers; and, the overemphasis on the negative aspects of programs in evaluation reports. Cox (1977) noted that evaluation research results are not utilized as fully as possible in program decision making processes and listed a number of possible reasons for this apparent underutilization: the mismatch between the roles and styles of program evaluators and agency personnel; the lack of experimental rigor found in most evaluation research designs which reduces their credibility; the failure of many evaluations to ask the correct program effectiveness questions; the overabundance of negative evaluation findings; the relative lack of specific program change recommendations made by evaluators; the lateness of many evaluation reports; and, the failure of many program evaluators to attend sufficiently to the dissemination of their evaluation's findings.

Attkisson and Broskowski (1978) observed that many human service providers viewed the process of program evaluation as critical, subjective, externally imposed and insensitive to the complexities and ambiguities of their programs. In addition, they were often not adequately prepared to participate in the process. Attkisson and Broskowski (1978) suggested that as a result, large scale national evaluation efforts were of limited utility and that evaluations should focus on local, limited issues.
Furthermore, evaluations should involve human service personnel from the outset in their conception and performance.

Wholey (1986) observed that there has been a significant increase in the utilization of evaluation results by state, local and not-for-profit organizations in the United States. In his view, timeliness and credibility of the evaluation findings are the key determinants of utilization. He further suggested that evaluations which produce ongoing feedback to programs as opposed to one-shot judgments of effectiveness are more likely to be utilized.

In sum, evaluation process characteristics thought to be critical in the determination of program evaluation utilization include degree of methodological rigor, credibility, evaluation practice style and the timeliness of results presentation.

The relatively sparse empirical research on this topic adds significantly to this picture by supporting some of these speculations and not others. For example, Knorr (1977) studied policy makers' use of social science research results in an Austrian setting through interviews with 70 government decision makers, and surveys of 628 Austrian social scientists. He found that most of the decision makers had made conceptual use of the social science research results but that desired instrumental uses were almost non-existent. In Knorr's study, research
process factors found to be related to the decreased likelihood of instrumental utilization centered around the cognitive and methodological inadequacy of most of the social science research projects studied. He noted that the results of most social science research studies by their very nature require further processing by decision and policy makers before they can be put to direct use.

Patton et al. (1978) studied the utilization of twenty evaluations of American federally-funded health programs through a series of structured interviews with program evaluators and decision makers. They found that conceptual utilization was common whereas instrumental utilization was not and that two major factors explained most of the variability in utilization: an evaluation process factor which they labelled the personal factor and a political considerations factor. In describing the personal factor, Patton et al. (1978) stated: "It is made up of equal parts leadership, interest, enthusiasm, determination, commitment, aggressiveness, and caring (on the part of program decision makers)" (p. 73). Further, this factor was brought into operation by evaluation processes that involved the program decision makers from the very beginning of the evaluation project in making major decisions with regard to how the evaluation would be conducted and utilized.
Among the factors unrelated to utilization was the methodological quality of the evaluation research design, a finding supported by Alkin et al.'s (1979) five case studies of evaluation utilization in U.S. education programs. As was the case with Patton et al. (1978), evaluators who involved key program decision makers in the planning, design and conduct of the evaluation were more likely to achieve conceptual utilization of the evaluation's findings.

Similarly, Weeks (1979) discovered that the methodological rigor of the evaluation's experimental design was inversely correlated with scores on the standardized index of program evaluation utilization that he developed and tested. In addition, he found that decision makers were more likely to utilize the results of evaluations that had employed multiple versus singular measures of program performance and collected data from a variety of sources instead of a single source. These and other findings led Weeks (1979) to conclude that "the greater the role of program personnel in shaping the evaluation, the more likely it is that the results will be used" (p. 149). In general, however, most of the correlations obtained in this study were very low with a multiple correlation between all the evaluation process variables and the utilization criterion measure of only .30, prompting the author to suggest that future research
should examine a greater variety of possible influences on evaluation utilization.

Finally, Greene (1988) also examined which factors influenced evaluation utilization in two case studies. She noted that the findings of the case studies "suggested a link between a participatory evaluation process and meaningful substantive uses of the evaluation results" (p. 100). She reported that early and frequent involvement of the program stakeholders in all phases of the evaluation process appeared to increase the likelihood of utilization success. In order to ensure maximum utilization, program stakeholders must be prepared by the evaluator to make use of the findings when they arrive.

Windle, Majchrzak and Flaherty (1979) examined nine years of program evaluation activity with the National Institute of Mental Health (NIMH) in the United States. During that time period, 50 studies of community mental health centers were conducted at a total cost of four million U.S. dollars. They noted that the bulk of the data produced by these studies was not used either instrumentally or conceptually to improve program operations. A factor related to this lack of utilization was the lack of evaluation processes designed to disseminate the information to the organization that was being evaluated or to other mental health centers. In this regard, they recommended that program evaluators build
dissemination activities into their plans for evaluation processes from the beginning stages of the evaluation and that larger scale evaluations be conducted, comparing the performances of a number of programs.

Weiss and Bucuvalas (1980) examined how social science research was being utilized by government officials and program managers and which factors affected the utilization potential of social science research results. Their research design involved a simulated situation in which actual government and program decision makers read reports from actual research studies, rated them on a number of qualities and then made judgments with regard to the utilization potential of the studies' research results. They found that five characteristics of social science research studies were correlated positively with the study's perceived usefulness: the quality of the research ($\bar{r} = .39$); the degree to which the research challenged the status quo ($\bar{r} = .19$); the degree to which the research results conformed with user expectations ($\bar{r} = .18$); the relevance of the research results to issues dealt with by the user ($\bar{r} = .15$); and finally, the amount of action orientation in the research ($\bar{r} = .12$). In total, all five of these research process variables accounted for approximately 45% of the variance in the judgments of usefulness made by these decision makers. Based upon these findings, these authors made a series of methodological
recommendations designed to increase the potential utility of social science research.

Siegel and Tuckel (1985) examined factors which influenced evaluation utilization in two evaluations of U.S. Community Mental Health Center programs. Evaluation process factors that were found to influence utilization included the nature of the evaluator's recommendations, the level of methodological sophistication of the evaluation research design, the timing of the presentation of the evaluation results, the amount of communication between the evaluator and the program decision makers and the manner in which the evaluator disseminated the findings.

When the speculative, non-empirical and empirical literature on evaluation process factors is combined, a clearer picture of their effects on program evaluation utilization emerges. This picture is presented in Table 2. Evaluation practice styles which foster the development of the "personal factor" identified by Patton et al. (1978) appear to enhance the likelihood of utilization. In addition, suggestions that the evaluation's credibility and timeliness influence evaluation utilization have also been confirmed in empirical studies. However, mixed evidence has been found as to the importance of methodological rigor as determinant of utilization. In general, the empirical research on evaluation process factors supports the view that multiple factors play roles in the determination of
Table 2

Evaluation Process Variables Related to Program Evaluation Utilization

<table>
<thead>
<tr>
<th>Variable</th>
<th>Author(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Methodological Sophistication*</td>
<td>Knorr (1977)</td>
</tr>
<tr>
<td></td>
<td>Siegel &amp; Tuckel (1985)</td>
</tr>
<tr>
<td>Decision Maker Takes Responsibility for Utilization*</td>
<td>Patton et al. (1978)</td>
</tr>
<tr>
<td>Participative Practice Style*</td>
<td>Patton et al. (1978)</td>
</tr>
<tr>
<td></td>
<td>Alkin et al. (1979)</td>
</tr>
<tr>
<td></td>
<td>Weeks (1979)</td>
</tr>
<tr>
<td></td>
<td>Greene (1988)</td>
</tr>
<tr>
<td>Multiple Measures Used</td>
<td>Weeks (1979)</td>
</tr>
<tr>
<td>Data from a Variety of Sources</td>
<td>Weeks (1979)</td>
</tr>
<tr>
<td>Focus on Utilization in Practice*</td>
<td>Windle et al. (1979)</td>
</tr>
<tr>
<td>Research Quality*</td>
<td>Weiss &amp; Bucuvalas (1980)</td>
</tr>
<tr>
<td>Challenge to Status Quo</td>
<td>Weiss &amp; Bucuvalas (1980)</td>
</tr>
<tr>
<td>Results Confirm Expectations</td>
<td>Weiss &amp; Bucuvalas (1980)</td>
</tr>
</tbody>
</table>
Table 2 (contd.)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Author(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Relevance to Decision Maker Needs*</td>
<td>Weiss &amp; Bucuvalas (1980)</td>
</tr>
<tr>
<td>Action Oriented Recommendations*</td>
<td>Weiss &amp; Bucuvalas (1980)</td>
</tr>
<tr>
<td>Timing of Results Presentation*</td>
<td>Siegal &amp; Tuckel (1985)</td>
</tr>
<tr>
<td>Communication Between Evaluator &amp; Decision Makers*</td>
<td>Siegal and Tuckel (1985)</td>
</tr>
<tr>
<td>Time on Results Presentation*</td>
<td>Siegal &amp; Tuckel (1985)</td>
</tr>
</tbody>
</table>

Note. In all cases the observed relationship between the evaluation process variable and utilization was positive unless noted otherwise in the text.

(*) The effects of these variables were examined in the current study.
program evaluation utilization.

Evaluator Characteristics

A small number of authors have hypothesized that program evaluator characteristics play a major role in determining the extent to which and manner in which evaluation findings will be utilized (Agarawala-Rogers, 1977; Brown & Braskamp, 1980; Caro, 1980; Ciarlo, 1981; Sadofsky, 1966; Weiss, 1966). These factors include the evaluator's level of understanding of the program, the evaluator's level of communication ability and the evaluator's organizational position in relation to the program being evaluated.

Sadofsky (1966) postulated that the evaluators' understanding of the program was a key factor in determining the extent to which evaluation results would be utilized. He argued that evaluators who did not clearly comprehend the goals and activities of a program would not likely produce a useful evaluation of a program. Similarly, Weiss (1966) noted that evaluators usually accept the descriptions of programs given by program staff as sufficient. She recommended that evaluators develop their own descriptions of programs and ensure that they possess a comprehensive understanding of program processes prior to
designing the evaluation. She also predicted that the evaluation findings produced by evaluators who followed there prescriptions would be more likely to be utilized.

Two evaluation theorists identified the evaluators' communication ability as a key determinant of utilization success. Agarawala-Rogers (1977) commented that evaluators were often not effective in communicating their findings to program decision makers and that this was an important influence on evaluation underutilization. She contended that this difficulty was exacerbated by the lack of specialists trained to act as liaisons between evaluators and decision makers. Caplan (1980) recommended that evaluators become more effective communicators and that they alter their communication style to one more suited to program decision makers.

Three other evaluation theorists contended that the organizational position of the program evaluator in relation to the program is a key determinant of program evaluation utilization. Ciarlo (1981) noted that evaluators who have a greater understanding of program processes because they are internal program staff are more likely to produce evaluation results that are utilized. Similarly, Caro (1980) argued that evaluation consultants who are external to the program being evaluated and have little or no authority are less likely than internal evaluators to produce evaluation results that will be utilized as
intended. Thus, he recommended that evaluators be attached to the regulatory bodies of programs and be given authority to ensure that their recommendations with regard to program change are implemented. Finally, Stevenson (1981) observed that the context and role demands faced by an internal evaluator are very different from those faced by external evaluators, and that these factors play a major role in determining utilization. He believed that internal evaluators should become aware of factors in their decision contexts such as their manager's style and financial pressures and incorporate this information into their evaluation planning and performance decisions. He predicted that internal evaluators who were able to develop these levels of awareness would be more likely to produce useful evaluation findings than external evaluators. However, Johnson (1980) found that the program evaluator's organizational position in relation to the organization being evaluated (internal versus external) was relatively unimportant in predicting utilization.

Brown and Braskamp (1980) contended that many different program evaluator characteristics such as gender, title, level of training and organizational position may play important roles in determining how credible their results are judged to be and indirectly, in the extent to which the results are utilized. Unfortunately, only a small number of empirical investigations of the influences of these and
other evaluator characteristics have been conducted, and this empirical research has failed to support the theoretical importance of understanding, communication ability and organizational position. However, a number of studies of simulated evaluation situations have suggested that a number of other evaluator characteristics (i.e., title, gender and extent of jargon use) may influence utilization (Table 3).

Braskamp, Brown and Newman (1982) reviewed a number of studies on the effects of various program evaluator characteristics on evaluation utilization potential using a persuasive communication model of the evaluation process. In these studies, the effects on evaluation utilization potential of the message sender (evaluator) characteristics were examined in simulated evaluation decision making situations. In one study (Braskamp, Brown, & Newman, 1978), the title of the evaluator was variously listed as researcher, evaluator or art specialist. Teachers and administrators rated the identical report credited to a "researcher" as significantly more objective than a report credited to someone with either one of the other titles. Presumably, a report judged as more objective would have higher utilization potential.

In a related study, Weeks (1979) had contradictory findings. He observed that the professional status of the program evaluator was unrelated to the level of utilization
Table 3

Evaluator Variables Related to Program Evaluation

<table>
<thead>
<tr>
<th>Utilization</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variable</td>
</tr>
<tr>
<td>Title*</td>
</tr>
<tr>
<td>Gender*</td>
</tr>
<tr>
<td>Use of Jargon</td>
</tr>
</tbody>
</table>

Note. (*) The effects of these variables on utilization were examined as part of the current study.
of the evaluation findings. He proposed that evaluators of higher professional status would be more likely to demand a greater role in program decision making processes subsequent to the evaluation, thereby increasing utilization of the findings.

In a second simulated study (Newman, Brown, & Litman, 1979), business and education professionals and students read identical evaluation reports which varied only in terms of the picture and name of evaluators of different genders placed on the front cover. Newman et al. (1979) found that if the report was about a business topic, the education subjects were more likely to agree with the recommendations of a male evaluator. Similarly, if the report was about an educational topic, the business subjects were more likely to agree with the recommendations of a male evaluator. Evaluator gender did not play a significant role when the report was in the subjects’ area of expertise.

In another study, Brown, Braskamp and Newman (1978) had educators read evaluation reports which were identical except for their use of jargon and the extent to which data were used to support the program evaluation recommendations. The educators rated jargon filled reports as more difficult to understand, a consideration which would presumably lead to decreased utilization.
In summary, a small number of empirical investigations of the effect of evaluator characteristics on utilization have suggested links between evaluator characteristics and utilization. The characteristics of gender and organizational position are included in the current study. The design of the current study, survey methodology, does not lend itself easily to the inclusion of other evaluator characteristics noted here including evaluator's communication ability, understanding of the program, use of jargon and title.

Decision Context Characteristics

Some evaluators and evaluation theorists have contended that characteristics of the decision context in which the program evaluation study is conducted play the major role in determining the extent and type of utilization achieved by an evaluation. The factor cited most often in these discussions is the failure of evaluators to understand and work with the complex, often bureaucratic decision making processes of various programs. Other factors that have been discussed as possible determinants of utilization include organizational resistance to change and the perception of evaluation as a threat by program staff, the presence of competing information in the decision context and public
reactions to evaluation findings.

Some degree of speculation has centred on the evaluator's failure to understand or appreciate the characteristics of the decision making processes in the program being evaluated. Thompson (1975) observed that evaluations are not always utilized as intended and that organizational and psychological phenomena such as administrative biases, personal conflicts and differences in the evaluator practice styles might account for this. He also contended that the utility of an evaluation was largely determined by its link with the decisions that it is designed to guide. Similarly, Guttentag (1977) argued that evaluators must come to appreciate the multiple perspectives of the different groups of program stakeholders if they wish to increase the potential utilization of their findings.

Caplan (1980) also observed that organizational forces play a major role in determining evaluation utilization success or failure. He noted that bureaucratic decision making systems are not well suited to the utilization of research results and recommended that the government experiment with the development of new types of organizational structures.

Much of the literature in this area has expounded the theme that evaluators have failed to understand the decision context in which they are operating. Dunn, Mitroff
and Deutsch (1981) noted that the failure of evaluators to correctly understand the decision context in which evaluation results are utilized has led to the obsolescence of evaluation research. They indicated that decision contexts are not entirely rational nor are their members anxiously awaiting the rational outputs produced by evaluation studies and therefore, recommended that evaluation theorists reconceptualize the evaluation problem in light of these misconceptions. This theme was echoed by De Young and Conner (1982) who observed that evaluators often hold the implicit assumption that program decision making follows a rational decision making model. Similarly, Saxe and Koretz (1982) noted that in a majority of cases, program evaluation research has not been useful (to the United States Congress) in the formation of public policy. They hypothesized that certain features of the decision context of the Congress such as the political nature of decision making play the most important role in determining patterns of utilization and argued that evaluators should become aware of these features and design their studies to meet the needs of congressmen and their aides.

Snapiro (1984) also contended that errors in the manner in which program evaluators have conceptualized organizational decision making systems are largely responsible for the failure of many evaluation studies being utilized. He noted that the decision context of
organizations is far more complex than had been believed by most evaluators and a variety of models of decision making are used within organizations under different conditions. Similarly, McLaughlin (1985) noted that most evaluators make incorrect assumptions with regard to the nature of organizational decision making processes. He noted that many evaluators have assumed that programs possess clear, well defined goals and are relatively stable in nature. In addition, many evaluators have assumed that the causal relations between program processes and clients are unidirectional. McLaughlin’s observations of how programs actually operate have led him to the view that programs are dynamic entities in a constant process of change and evolution. He suggested that evaluators must be aware of these facts and should incorporate this awareness into their evaluation processes if they wish to achieve greater utilization.

Soloman and Shortell (1982) believed that some problems that reduce evaluation utilization are organizational rather than methodological in nature. They noted that the information needs of decision makers are different from those of evaluators and that one of the key factors in determining utilization is the amount of commitment of the organizational decision makers to the evaluation activity.
Stahler and Tash (1982) hypothesized that evaluations are not fully utilized as intended because evaluators have been insensitive to the unique characteristics of the programs that they have been asked to evaluate. They observed that in the context of mental health programs, program evaluation is not an end in itself but rather one important part of the management information and program development systems. Therefore, they argued that evaluators who become aware of and use the unique features of mental health programs in the design of their evaluations will likely achieve greater utilization.

Finally, Myers and Gonda (1987) contended that the likelihood that an evaluation's findings will be utilized can be increased by considering some of the features of the decision making context in which the evaluation is being conducted. They indicated that program evaluation processes should be adapted to meet the needs of the organization and that evaluators should respond to the information needs of key decision makers within these organizations.

Some evaluation theorists have argued that the likelihood of program evaluation utilization is reduced because program staff perceive evaluation as a threat. For example, Agarawala-Rogers (1977) contended that program personnel sometimes view program evaluation as a tool used by government funders to reduce or eliminate the funding of their programs. In addition, Wilderman (1980) noted that
because of a perception of evaluation as not often useful, program managers and line staff can be expected to react to the introduction of evaluation with both fear and hostility because it diverts valuable staff resources from direct service and wastes scarce monetary resources.

Suchman (1967) indicated that the forces at work within the organization being evaluated, such as a high degree of resistance to change, would influence the extent of program evaluation utilization. Thus, he suggested that program decision makers and staff are personally invested in program structures and processes and will resist efforts to modify them as a result of program evaluation findings.

Some authors have cited a number of other individual decision context factors as primary determinants of program evaluation utilization success. For example, Schulberg and Baker (1971) hypothesized that one key factor was the administrator's motivation in initiating the evaluation. For example, some decision makers initiated evaluations to improve programs whereas others initiated evaluations in attempts to attack or terminate programs. They also contended that inadequate utilization of evaluation findings could be the result of blocked feedback.

In addition, Cumming (1976) noted that the evaluations of American Community Mental Health Center programs are not often utilized as intended. He indicated that much of this apparent non-utilization was the result of serious
misunderstandings with regard to the role of evaluation research in the evaluation of mental health program effectiveness. Based upon his experiences in evaluating one Canadian mental health centre, he indicated that mental health centre personnel usually perceive that evaluations are conducted for the primary purpose of quality assurance although judgments about the effectiveness of program elements are what are really needed. Cox (1977) noted that evaluation findings are only one of many sources of information that program decision makers use in the program development process. He hypothesized that the presence of other, more powerful sources of competing information might reduce the utilization of an evaluation study's findings.

Stevenson and Ciarlo (1982) predicted that more program evaluation utilization in mental health programs would occur at local than state levels because of differences in the organizational contexts of the two settings. Specifically, they noted that local decision makers usually have a more immediate concern for their programs than do state officials and as a result are more prepared to utilize the results of program evaluations. In addition, decisions made at the local level are not as dominated by financial issues as those made at the state level. Furthermore, locally-based evaluators are likely to have a greater understanding of local decision context issues, resulting in a greater likelihood for the
production of potentially useful evaluation designs and results.

Finally, Manga (1987) observed that the lack of clear, measurable objectives within some programs creates a very difficult situation for evaluators and decreases the likelihood that their results will be utilized. He also noted that the resistance to change within organizations and the complex political and value systems in which organizations are imbedded can adversely affect the potential utilization of an evaluation's results.

Most empirical investigations of factors believed to be related to program evaluation utilization have examined the effects of decision context factors. Many of the theoretical speculations in this area have been supported by the results of these investigations. These findings are summarized in Table 4.

Bigelow and Ciarlo (1976) found that program evaluation utilization in a community mental health center program was largely influenced by factors specific to the decision context of mental health centers. They noted that evaluation findings were not the only input into the management decision context in these settings; decision context factors such as staff input and staff capability played much more of a role in influencing decision outcomes than did evaluation findings.
Table 4

Decision Context Variables Related to Program Evaluation

Utilization

<table>
<thead>
<tr>
<th>Variable</th>
<th>Author(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Presence of Competing Information*</td>
<td>Bigelow &amp; Ciarlo (1976)</td>
</tr>
<tr>
<td>Differences in Decision Maker &amp; Evaluator Information Needs</td>
<td>Knorr (1977)</td>
</tr>
<tr>
<td>Political Considerations</td>
<td>Patton et al. (1978)</td>
</tr>
<tr>
<td>Differences in Decision Maker &amp; Evaluator Decision Making Styles</td>
<td>Alkin et al. (1979)</td>
</tr>
<tr>
<td>Ambiguity in Program Goals</td>
<td>Weeks (1979)</td>
</tr>
<tr>
<td>Number of Decision Makers*</td>
<td>Weeks (1979)</td>
</tr>
<tr>
<td>Management Support Scale*</td>
<td>Weeks (1979)</td>
</tr>
<tr>
<td>Need to Evaluate Scale*</td>
<td>Weeks (1979)</td>
</tr>
<tr>
<td>Variable</td>
<td>Author(s)</td>
</tr>
<tr>
<td>-----------------------------------------</td>
<td>-------------------------------</td>
</tr>
<tr>
<td>Presence of Information Brokers</td>
<td>Johnson (1980)</td>
</tr>
<tr>
<td>Turnover of Key Staff Members</td>
<td>Bedell et al. (1985)</td>
</tr>
<tr>
<td>Organizational Resistance to Change</td>
<td>Siegel &amp; Tuckel (1985)</td>
</tr>
</tbody>
</table>

**Note.** (*) The effects of these variables on utilization were examined as part of the current study.
Knorr (1977) surveyed European government decision and policy makers with regard to the extent of their utilization of social science research findings. Based on his findings he reported that these government-based decision makers have concrete information demands that are not often met by the outputs from social science research studies and that this shortcoming had led to deceased instrumental utilization of social science research results in Austria.

Patton et al. (1978) reported on the amount of utilization achieved in a sample of twenty American health program evaluations. Political considerations in the decision context that affected utilization included interagency and intra-agency rivalries, budgetary issues, power struggles and internal debates. Patton et al. found that these issues surfaced during the ongoing processes of the evaluation and recommended that evaluators anticipate the potential effects of these factors prior to entry into these programs and use this forewarning in planning evaluation processes.

Alkin et al. (1979) reported that organizational factors such as an administrator's decision making style and the orientation of potential users to the evaluation process played some role in determining the utilization achieved by a sample of five evaluations of American educational programs. Weeks (1979) found that programs
rated as having greater levels of ambiguity in program goals were less likely to utilize evaluation results. In addition, it was observed that, surprisingly, when a greater number of decision makers were involved, the evaluation results were more likely to be utilized. Thus, he suggested that this finding may be a reflection of the importance assigned to the program evaluation by organizational decision makers. Program evaluations with more decision makers involved were those in which there was a greater degree of interest in the outcomes. He also found that the amount of management support and need to evaluate expressed by program decision makers prior to the start of the evaluation were related to utilization.

Johnson (1980) found that the presence of information brokers within the program decision context was related to increased utilization of the evaluation findings. The proximity of the decision makers to the programs being evaluated was also related to increased utilization. Finally, the amount of contact and involvement of the decision makers in the evaluation processes was also related to greater utilization. Evaluations on which program decision makers spent more time and effort were more likely to produce findings which were utilized.

Bedell et al. (1985) reported that staff changes within mental health organizations were related to decreased utilization of evaluation results. When a key
staff member left the program before the recommendations for program change derived from an evaluation could be implemented or when key staff were given changes in their duties or assignments, the results were less likely to be utilized.

Finally, Siegel and Tuckel (1985) revealed that organizational resistance to change could reduce the utilization of evaluation results. They viewed this resistance as an inevitable property of organizations that evaluators had to work to neutralize if they hoped for higher levels of utilization.

In summary, both speculative and empirical literature point to the importance of decision context characteristics in determining evaluation utilization. A number of empirical investigations have supported the view that characteristics of the decision contexts in which program evaluations are conducted directly influence utilization. Factors such as the presence of competing information, the level of management support for the evaluation and political considerations can play a role in determining the extent of evaluation utilization.
Problem Statement

As a result of the perception that program evaluation results are not often utilized, the future of program evaluation may be in jeopardy (Brown & Braskamp, 1980; Ciarlo, 1981; Conner, Altman, & Jackson, 1984; Cook, 1984; Cronbach et al., 1980; Dunn, Mitroff, & Deutsch, 1981; Patton, 1986; Shadish & Reichardt, 1987; Weiss, 1977). If this perception is not changed, potential funders of program evaluation may choose to reduce the financial resources available to conduct evaluations. In addition, current evaluators may leave the field and prospective evaluators may decide to specialize in other, more viable areas.

Ironically, the concern that evaluation research is underutilized is based for the most part on (soft) anecdotal reports and subjective perceptions of evaluators and program decision makers rather than on hard empirical data. For research on evaluation utilization did not begin until the early 1970s and was not systematically reviewed until the 1980s. As a result, the accuracy of the perception that evaluation studies are underutilized is not clear at all. Data that directly address the question of accuracy have been collected as part of the current study.

This uncertainty (regarding the actual extent of program evaluation utilization) is exacerbated for Canadian
program evaluators by the simple fact that most of the past research on evaluation utilization has been conducted in the United States. For instance, research has demonstrated that the organizational context in which an evaluation is conducted can influence utilization, and the Canadian organizational context for evaluation differs in some important ways from the U.S. context (Mayne, 1986; Rutman, 1986; Rutman & Mayne, 1985; Wholey, 1986). Another important difference is the Canadian provinces/federal government relationship versus the American states/federal government relationship. In addition, the mandated evaluation function within the Canadian federal government which ensures that all Canadian federal departments have a program evaluation department with certain standard features contrasts sharply with the American situation in which the evaluation capability varies widely among federal government departments. Finally, the nationally implemented mental health and education programs of the U.S. have no equivalents in the provincially dominated education, health and mental health systems in Canada. Therefore, American research findings on evaluation utilization may not be directly generalizable to the Canadian context.

The current study provides the first systematic collection of data on Canadian program evaluation utilization. In addition, it is designed, in part, to replicate a number of the more prominent American
investigations.

Another problem with the current understanding of evaluation utilization is that most previous research investigations have utilized only program decision makers as subjects. It is difficult, therefore, to determine the extent to which evaluators and program decision makers share perceptions of the precise processes of program evaluation. This is a potential shortcoming which is addressed by the present study in which a large sample of evaluators will serve as the respondents.

A final problem with previous research on evaluation utilization is that earlier studies used loosely structured and highly subjective interview or survey questions of unknown reliability and validity as dependent variables (Conner, 1981; Cousins & Leithwood, 1986; Dunn, 1983; Weiss, 1982). Despite the existence of utilization scales of demonstrated reliability and validity (e.g., Johnson, 1980; Weeks, 1979), these measures have been used as dependent variables in only a small minority of evaluation utilization studies. As a result, the findings of most studies of program evaluation utilization must be viewed and compared with caution. Both of these evaluation utilization scales have been employed in the current study.
Research Objectives and Hypotheses

The research objectives and main hypotheses of the current study are as follows:

1) To determine the extent to which three types of program evaluation utilization are to be found in a recent and representative sample of Canadian program evaluations. It is hypothesized that:

\[ H_0: \quad \text{The probability of occurrence of program evaluation utilization will be zero (} \mu_{PEU} = 0). \]

\[ H_A: \quad \text{The probability of occurrence of program evaluation utilization will be significantly greater than zero (} \mu_{PEU} > 0). \]

\[ H_0: \quad \text{The probabilities of occurrence of the three types of utilization (conceptual, instrumental and persuasive) will be equal (} \mu_{CPEU} = \mu_{IPEU} = \mu_{PPEU}). \]

\[ H_A: \quad \text{The probability of occurrence of conceptual evaluation utilization will be significantly greater than that of the other two types (} \mu_{CPEU} > \mu_{IPEU} \text{ or } \mu_{PPEU}). \]

2) To examine the relationships between: selected characteristics of a sample of recent Canadian program evaluations, characteristics of the program evaluators who conducted them and characteristics of the decision context
in which the evaluations were conducted, and the type of program evaluation utilization (instrumental, conceptual and persuasive). It is hypothesized that:

\[ H_0 \] Characteristics of the evaluation process, evaluator and decision context will be unrelated to the occurrence or non-occurrence of program evaluation utilization ($\chi^2$ and $\hat{f} = 0$).

\[ H_a \] Characteristics of the evaluation process, evaluator and decision context will be significantly related to the occurrence or non-occurrence of evaluation utilization ($\chi^2$, and $\hat{f}$ are significantly greater than 0).

3) To determine if these selected program evaluation process, program evaluator and decision context variables will account for a significant portion of the variance in the extent of utilization success and to determine which of these groups of variables and individual variables are most highly related to utilization. It is hypothesized that:

\[ H_0 \] No one of the groups of evaluation process, evaluator or decision context variables will be more highly related to the occurrence or non-occurrence of utilization than any other.

\[ H_a \] Variables from the group of evaluation process variables will be more highly related to the occurrence and non-occurrence of utilization than variables from the other two groups.
4) To test three specific hypotheses about the relationship of specific evaluation process and evaluator variables to program evaluation utilization that have been suggested by past studies of program evaluation utilization in the United States to determine whether they can be replicated in the Canadian context. These hypotheses are:

H5: $H_0$: The evaluation practice style of the evaluator will be unrelated to the occurrence or non-occurrence of conceptual and instrumental utilization ($\chi^2 = 0$).

$H_A$: The stakeholder-oriented evaluation practice style will be associated with greater probabilities of the occurrence of both conceptual and instrumental utilization (from Patton et al., 1978; Alkin et al., 1979; Greene, 1988).

H6: $H_0$: Evaluators who are internal to the program being evaluated will not achieve a greater probability of instrumental utilization than those program evaluators who are external to the program ($\chi^2 = 0$).

$H_A$: Evaluators who are internal to the program will achieve a greater probability of instrumental utilization than those who are external to the program (from Bigelow & Ciarlo, 1976; Johnson, 1980; Stevenson & Ciarlo, 1982).

H7: $H_0$: The presence of a program decision maker who took personal responsibility for ensuring program evaluation utilization will be unrelated to greater probabilities of instrumental and conceptual utilization ($\chi^2 = 0$).
The presence of a program decision maker who took direct responsibility for ensuring utilization will be related to higher probabilities of both instrumental and conceptual utilization (from Patton et al., 1978).
CHAPTER II

METHOD

Subjects

The subjects in the present study were 332 program evaluators who had recently conducted a Canadian program evaluation study and who responded voluntarily to a mail-out survey. Subjects included members of Canada's two principle program evaluation organizations, the Canadian Evaluation Society (CES) and the Canadian Psychological Association's Program Evaluation Section (CPA-PES), as well as members of Canadian university departments who were not members of either organization. Of the 332 subjects, 260 were members of the CES only, 17 belonged to the CPA-PES only and 15 were members of both program evaluation organizations. The remaining 40 subjects were not evaluation organization members.

Data Collection Procedure

In order to collect data on program evaluation utilization from a representative sample of program evaluators who had recently conducted Canadian program evaluations, two separate data collection methods were
used, the Total Design Method (Dillman, 1978) and a university mailing method.

**Total Design Method**

The names and addresses of CES and CPA-PES members were obtained from the membership mailing lists of the CES and the CPA Program Evaluation Section which were last updated on January 1, 1988. Then, following Dillman’s (1978) prescription, the 868 CES and CPA-PES members were sent a personally addressed mailed survey package on September 20, 1988. The survey package contained a personalized cover letter in either French or English (Appendix A), a professionally designed and printed survey (Appendix B) which included a human subject consent form and a postage-paid return envelope addressed to the researcher. One week later, on September 27, a bilingual reminder card (Appendix C) was sent to all CES and CPA-PES members. On October 25th, 1988, all of the CES and PES-CPA members from whom a response of some type had not been received were sent a second survey package which included a revised cover letter in either French or English (Appendix D), a second copy of the survey, and a second postage-paid and pre-addressed return envelope.
Surveys were identified with a subject code number which was linked to the name and address of each CES and CPA-PES member. This procedure was followed to enable the researcher to identify individual CES and CPA-PES respondents so that their names and addresses could be removed from the mailing list, thereby avoiding a second mailing of the survey to subjects who had already responded.

**University Mailing Method**

With the second method, an attempt was made to contact program evaluators who were not members of either CES or CPA-PES by mailing 431 non-identified survey packages (also on September 20, 1988) to the department heads of 135 Canadian university departments of Criminology, Education and Educational Administration, Public Administration, Psychology, Social Work and Sociology. See Appendix E for both French and English versions of the cover letters sent to the university department heads.

The university department heads were asked to identify members of their departments who might have recently completed a Canadian program evaluation study and to supply these individuals with a non-identifying survey package. These survey packages contained a non-identifying cover
letter in either English or French (Appendix F), a numbered
survey and a postage-paid return envelope addressed to the
researcher. The potential study participants who were based
in Canadian university departments were not subject to any
of the follow-up procedures included as part of the Total
Design Method.

Program Evaluation Utilization Survey

The measurement instrument used in the present study
was a survey (Appendix B) which included: a series of
screening questions, a group of program evaluation
utilization measures (dependent variables) and measures of
three groups of evaluation process, evaluator and decision
context variables (independent variables). It was
constructed following procedures prescribed by Dillman
(1978). An earlier version of the survey was pilot tested
during July, 1988, with six members of the Psychology
department at the University of Windsor who had conducted
Canadian program evaluation studies at some time in the
past.
Screening Questions

On the first page of the survey, potential study participants were informed about the nature and importance of the study and the identity of the researcher. They were also assured of confidentiality. They were then asked to return the survey in the pre-paid, addressed envelope even if they did not wish to participate in the study in order to ensure that they would not be recontacted with follow-up mailouts.

On the second page of the survey, potential study participants were asked to read and sign a consent form which detailed the purpose of the study. In addition, they could request a summary of the study results by providing a current mailing address. Potential study participants were told that the consent form which contained their identification number would be removed from the survey upon receipt by the researcher after their name had been removed from the list of non-respondents.

Potential study participants who were willing to complete the survey were then screened further using three questions on the third page of the survey. Survey respondents were first asked if they had participated as evaluators in a Canadian program evaluation project completed between January 1, 1985 and January 1, 1988. If they had not, they were asked to return the survey without
answering any more questions. This screening question had two purposes. First, since not all CBS and CPA-PES or university department members identified by their department head were program evaluators, this question eliminated them from participation in the study at the outset. Second, it was believed that restricting responses to the three-year time period immediately preceding the beginning of 1988 would reduce the number of errors that subjects might make in recalling details of evaluation projects and also would have allowed sufficient time for the utilization of evaluation results to occur.

The remaining potential study participants then indicated their role in the last evaluation study that they had conducted or participated in between January 1, 1985 and January 1, 1988: Principal Investigator, Evaluation Project Director, Research Assistant or Other (specify). After answering this question, the subjects were then instructed to complete the remainder of the survey only if they believed that they had detailed knowledge of the evaluation project.

Potential study participants were then asked if they were aware of how the results of the last Canadian program evaluation project that they completed between January 1, 1985 and January 1, 1988 had been utilized. The purpose of this screening question was to eliminate those evaluators who could not complete the critical measures of program
evaluation utilization success and to focus the study participants on a single Canadian program evaluation that had been completed recently.

Survey respondents who had not been eliminated by any of these questions were then asked to complete the remainder of the survey.

**Dependent Variables**

The dependent variables were chosen either on the basis of their demonstrated reliability and validity or because of the importance ascribed to them in the evaluation utilization literature. They included two program evaluation utilization scales (Johnson, 1980; Weeks, 1979), and three dichotomous indicators of subject judgment with regard to the occurrence or non-occurrence of three different program evaluation utilization types (instrumental, conceptual and persuasive).

The Weeks (1979) Program Evaluation Utilization Scale consists of ten items that reflect different extents and types of program evaluation utilization (Appendix B, pp. 4-5). Subjects were instructed to select the one statement of the ten which best described the manner in which the results of their program evaluation study had been utilized. This procedure yielded a score between 1 and 10.
Weeks (1979) reported a test retest correlation of $r = .84$ as an indication of this scale's reliability. In addition, Weeks found significant correlations between a number of evaluation process and decision context variables and the score on this scale.

The Johnson (1980) Program Evaluation Utilization Scale consists of seven items that, like Weeks' scale, reflect different extents and types of program evaluation utilization (Appendix B, pp. 5-6). Subjects were instructed to indicate whether or not each of the seven statements was true of their program evaluation study. This procedure yielded a total score between 1 and 7. Johnson (1980) reported a coefficient of reproducibility of $r = .94$ and a coefficient of scalability of $\alpha = .73$ as indices of the reliability of this scale. Johnson also found that the score on this scale was significantly correlated with a number of program evaluation process and decision context characteristics.

After completing these two scales, subjects were asked to indicate which type or types of program evaluation utilization (if any) had occurred in the evaluation in question. The choices were: No Utilization, Instrumental Utilization, Conceptual Utilization, Persuasive Utilization and Other Type of Utilization. Each of the three identified types of utilization was described to facilitate consistent use of the terms by the subjects. It should be stressed at
this point that these three categories were non-exclusive. In other words, it was expected that some evaluations would be described as belonging in only one of the utilization categories, some would be assigned to two categories and, in some cases, the same evaluation would be assigned to all three utilization categories.

**Independent Variables**

The remainder of the questionnaire consisted of a series of questions designed to obtain information on evaluation process, evaluator and decision context characteristics of the evaluation in question. Specific independent variables were selected for inclusion in each of these three categories because of their demonstrated reliability and validity in past program evaluation utilization studies in which they were significantly related to utilization of some type and/or their speculated influence on utilization in the relevant literature. The relationships among the evaluation process, evaluator and decision context variables chosen for inclusion in the current study and the literature on these issues are summarized in Tables 5, 6 and 7.

The order in which subjects answered questions, with one or two exceptions, was: decision context, evaluator and
Table 5

Relationships Between Evaluation Process Variables and Program Evaluation Utilization (PEU)

<table>
<thead>
<tr>
<th>VARIABLE</th>
<th>CITATIONS</th>
<th>RELATIONSHIP TO PEU</th>
</tr>
</thead>
<tbody>
<tr>
<td>Primary Decision Maker's Amount of Involvement in the Process</td>
<td>Patton et al. (1978)</td>
<td>+</td>
</tr>
<tr>
<td>Evaluation Activity Focus</td>
<td>Cox (1977); Attkisson &amp; Broskowski (1978); Cronbach (1982); Rutman (1986)</td>
<td>+/-</td>
</tr>
<tr>
<td>Relevance of Findings</td>
<td>Weiss &amp; Bucuvalas (1980)</td>
<td>+</td>
</tr>
<tr>
<td>Methodological Rigor</td>
<td>Rossi (1966); Suchman (1967); Berk &amp; Rossi (1977); Cox (1977); Knorr (1977); Patton et al. (1978); Alkin et al. (1979); Levitan &amp; Wurzberg (1979); Weeks (1979); Weiss &amp; Bucuvalas (1980)</td>
<td>+; Unrelated; +; Unrelated; +; -</td>
</tr>
</tbody>
</table>


Table 5 (contd).

<table>
<thead>
<tr>
<th>VARIABLE</th>
<th>CITATION(S)</th>
<th>RELATIONSHIP TO PEU</th>
</tr>
</thead>
<tbody>
<tr>
<td>Evaluation Practice Style</td>
<td>Davis &amp; Salasin (1975);</td>
<td>+/-</td>
</tr>
<tr>
<td></td>
<td>Patton (1978):</td>
<td>+/-</td>
</tr>
<tr>
<td></td>
<td>Patton et al. (1978);</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>Alkin et al. (1979);</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>Neigher (1979);</td>
<td>+/-</td>
</tr>
<tr>
<td></td>
<td>Rich (1979);</td>
<td>+/-</td>
</tr>
<tr>
<td></td>
<td>Weeks (1979);</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>Rutman (1980);</td>
<td>+/-</td>
</tr>
<tr>
<td></td>
<td>Cronbach (1982);</td>
<td>+/-</td>
</tr>
<tr>
<td></td>
<td>Heilman (1982);</td>
<td>+/-</td>
</tr>
<tr>
<td></td>
<td>Rafter (1984);</td>
<td>+/-</td>
</tr>
<tr>
<td></td>
<td>Rog &amp; Bickman (1985);</td>
<td>+/-</td>
</tr>
<tr>
<td></td>
<td>Pancer (1985);</td>
<td>+/-</td>
</tr>
<tr>
<td></td>
<td>Greene (1988)</td>
<td>+</td>
</tr>
<tr>
<td>Evaluation Findings</td>
<td>Weiss (1966);</td>
<td>+/-</td>
</tr>
<tr>
<td></td>
<td>Agarawala-Rogers (1977);</td>
<td>+/-</td>
</tr>
<tr>
<td></td>
<td>Patton et al. (1978)</td>
<td>Unrelated</td>
</tr>
<tr>
<td>Relevance to Decision Maker's Needs</td>
<td>Weiss &amp; Bucuvalas (1980)</td>
<td>+</td>
</tr>
<tr>
<td>Timeliness</td>
<td>Guttentag (1977);</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>Patton et al. (1978);</td>
<td>Unrelated</td>
</tr>
<tr>
<td></td>
<td>Siegal &amp; Tuckel (1985);</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>Rayner (1986)</td>
<td>+</td>
</tr>
</tbody>
</table>
### Table 5 (contd).

<table>
<thead>
<tr>
<th>VARIABLE</th>
<th>CITATION(S)</th>
<th>RELATIONSHIP TO PEU</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time Since Evaluation</td>
<td>Caplan (1980)</td>
<td>+</td>
</tr>
<tr>
<td>Final Report Made</td>
<td>Agarawala-Rogers (1977);</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>Windle et al.</td>
<td>+</td>
</tr>
<tr>
<td>Recommendations Made</td>
<td>Weiss &amp; Bucuvalas (1980)</td>
<td>+</td>
</tr>
<tr>
<td>Budget</td>
<td>Patton et al. (1978)</td>
<td>Unrelated</td>
</tr>
</tbody>
</table>
Table 6

**Relationships Between Evaluator Variables and Program Evaluation Utilization (PEU)**

<table>
<thead>
<tr>
<th>VARIABLE</th>
<th>CITATION(S)</th>
<th>RELATIONSHIP TO PEU</th>
</tr>
</thead>
<tbody>
<tr>
<td>Evaluation</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>Specialization</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>Primary Work</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>Setting</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Internal vrs External to Program</td>
<td>Bigelow &amp; Ciarlo (1976); Johnson (1980)</td>
<td>+ Unrelated</td>
</tr>
<tr>
<td>Organizational Position</td>
<td>Caro (1980); Ciarlo (1981); Stevenson (1981)</td>
<td>+/- +/- +/-</td>
</tr>
<tr>
<td>Age</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>Gender</td>
<td>Newman et al. (1978)</td>
<td>+/-</td>
</tr>
<tr>
<td>Educational Level</td>
<td>Weeks (1979)</td>
<td>Unrelated</td>
</tr>
<tr>
<td>Academic Discipline</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>Current Discipline</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>Evaluator Experience</td>
<td>*</td>
<td>*</td>
</tr>
</tbody>
</table>
Table 6 (contd.)

<table>
<thead>
<tr>
<th>VARIABLE</th>
<th>CITATION(S)</th>
<th>RELATIONSHIP TO PEU</th>
</tr>
</thead>
<tbody>
<tr>
<td>Practice Influences</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>Utilization Enhancing Behaviors</td>
<td>Braskamp &amp; Brown (1980)</td>
<td>+</td>
</tr>
</tbody>
</table>

Note. (*) Variables in which no citation or relationship to program evaluation utilization are indicated were chosen based upon the current author's own evaluation experience and/or upon the anecdotal reports of other evaluators.
Table 7

Relationships Between Decision Context Variables and Program Evaluation Utilization (PEU)

<table>
<thead>
<tr>
<th>VARIABLE</th>
<th>CITATION(S)</th>
<th>RELATIONSHIP TO PEU</th>
</tr>
</thead>
<tbody>
<tr>
<td>Program Type</td>
<td>Stahler &amp; Tash (1982);</td>
<td>+/-</td>
</tr>
<tr>
<td></td>
<td>Stevenson &amp; Ciarlo (1982)</td>
<td>+/−</td>
</tr>
<tr>
<td>Program Tenure</td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>Program Size</td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>Program Decision Making Centre</td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>Primary Decision Maker’s Tenure</td>
<td>Bedell et al. (1982)</td>
<td>+</td>
</tr>
<tr>
<td>Number of Decision Makers</td>
<td>Guttentag (1977); Weeks (1979)</td>
<td>−  +</td>
</tr>
<tr>
<td>Evaluation Initiation Source</td>
<td>Schulberg &amp; Baker (1971); Johnson (1980)</td>
<td>+/−</td>
</tr>
<tr>
<td>Evaluation Funding Source</td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>Number of Past Evaluations</td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>Past Evaluation Findings</td>
<td>*</td>
<td></td>
</tr>
</tbody>
</table>
Table 7 (contd.)

<table>
<thead>
<tr>
<th>VARIABLE</th>
<th>CITATION(S)</th>
<th>RELATIONSHIP TO PEU</th>
</tr>
</thead>
<tbody>
<tr>
<td>Organizational Constraints</td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>Competing Information</td>
<td>Bigelow &amp; Ciarlo;</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Cox (1977);</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Shapiro (1984)</td>
<td>-</td>
</tr>
<tr>
<td>Management Support</td>
<td>Weeks (1979)</td>
<td>+</td>
</tr>
<tr>
<td>Evaluation Need</td>
<td>Weeks (1979)</td>
<td>+</td>
</tr>
</tbody>
</table>

Note. (*) These variables represent attempts to operationalize features of the decision context believed by many evaluation theorists to play major roles in the determination of program evaluation utilization.
evaluation process characteristics. This order was deliberately chosen by the investigator such that potentially sensitive questions were presented later in the survey. For the sake of continuity with previous discussions, the order in which they will be discussed here is: evaluation process, evaluator, and decision context characteristics.

**Evaluation Process Characteristics**

**Primary Decision Maker's Amount of Involvement in the Evaluation Process:** Number of contact hours spent with the evaluator during planning, conduct and utilization phases of evaluation.

**Evaluation Activity Focus:** Percentage of total evaluator time spent on Evaluability Assessment, Process Evaluation, Outcome Evaluation, Dissemination of Evaluation Findings and Other Evaluation Activities.

**Relevance of Evaluation Findings:** Evaluator’s 5-point Likert type rating of relevance of evaluation findings for decision makers’ needs.
Degree of Methodological Rigor in the Evaluation Design:
Three ordered categories of evaluation method rigor: true experiment (most rigorous), quasi-experiment, other quantitative or qualitative methods (least rigorous).

Evaluation Findings: Overall judgement of program effectiveness made as result of evaluation: five-point Likert scale from 0 (Not at all effective) to 4 (Totally effective).

Timeliness of Evaluation Findings: Series of four ranked categories from "On time" to "Too late to be utilized".

Time Since the Evaluation Was Completed: Amount of time in months between completion of questionnaire and evaluation. This is amount of time available for program evaluation utilization to have occurred.

Presence or Absence of a Written Final Report: Whether or not written final evaluation report was submitted to program decision makers at completion.

Presence or Absence of Recommendations by the Evaluator in the Final Report: Whether or not evaluator provided recommendations for program change based on evaluation findings in final report.
Evaluation Budget: Total amount of money spent on evaluation during its entire period of operation.

Evaluator Characteristics

Degree of Specialization in Evaluation: Percentage of total work time subject spent in planning and conducting program evaluations.

Evaluator’s Primary Evaluation Work Setting: Work setting in which the evaluator spent most of his/her time conducting program evaluations.

Internal versus External Evaluator: Whether or not evaluator was a member of evaluated program.

Evaluator’s Organizational Position: Evaluator’s organizational position in relation to the evaluated program. A series of 7 possibilities including line staff temporarily assigned to evaluation duties, evaluation unit staff member and outside consultant.

Evaluator’s Age: Number of years since evaluator’s year of birth expressed as whole number.
Evaluator's Gender: Female or Male.

Effect of Evaluator's Gender on Evaluation Practice: Five-point Likert Scale from Very Negative Effect to Very Positive Effect.

Evaluator's Highest University Degree: No degree, Bachelor's Degree, Master's Degree or Doctorate.

Academic Discipline of Evaluator's Highest University Degree: A forced choice among 16 academic disciplines some evaluators are known to come from and a category of other.

Evaluator's Current Disciplinary Affiliation: A forced choice among same 17 categories as previous question.

Evaluator's Amount of Program Evaluation Experience: Number of program evaluations conducted by subject including evaluation in question.

Evaluator's Similarity to Evaluation Practice Patterns: Five-point Likert type ratings of evaluator's self-perceived similarity to four different evaluation practice styles: academic, local change agent, stakeholder-oriented and effectiveness judge (adapted from Shadish & Epstein, 1987).
Evaluator's Primary Practice Style: The one pattern of the four listed above that the evaluator judged to be most similar to own evaluation style in the evaluation in question.

Evaluation Practice Influences: Influences on evaluation practice, if any, cited by evaluator.

Utilization-Enhancing Behavioral Checklist: A checklist of sixteen behaviors thought to promote the utilization of program evaluation findings adapted from Brown and Braskamp (1980). The total score out of a possible sixteen was determined and entered into the data analyses.

Decision Context Characteristics

Type of Program: Federal government, provincial government, not-for-profit, for-profit or other type of program.

Program Tenure: Length of time (in years) that evaluated program had been in operation.

Program Size: Total number of staff employed by and volunteering with the program (in six ordered categories).
Program Decision Making Centre: Location of major program
decision making activities (local level, local parent body,
non-local parent body or other).

Primary Decision Maker's Tenure: Length of time (in years)
that program's primary decision maker had held her/his
position.

Source of Evaluation Initiation: Program personnel,
program's parent body, program's funder or other overseeing
body and other.

Source of Evaluation Funding: Same four forced choices as
previous question.

Number of Previous Evaluations of the Program: A whole
number that potentially ranged from 0 to infinity.

Valence of Past Evaluation Findings: A five-point rating of
program's general effectiveness as judged by any past
evaluations.

Organizational Constraints on the Evaluation: Three five-
point ratings of number of organizational constraints
experienced by the evaluator during planning,
implementation and dissemination phases.
Organizational Decision Making Complexity: Number of program decision makers who had to be convinced by evaluator to utilize evaluation findings.

Competing Information: Presence or absence or information in decision context that competed with evaluation findings in decision making process.

Management Support: From Weeks (1979), a sub-scale designed to assess the amount of management support for the evaluation of the program. Based on two items on which the subjects were asked to supply ratings from 0 to 3. Items were "Did you feel that most program decision makers in this program and its parent body supported this evaluation of their program?" and "Do you think that most of the key decision makers in this program and its parent body will use this evaluation's results as a guide to making program improvements?". The scale score is made up of the total of the ratings on both items and could therefore range from 0 to 6. Weeks (1979) found that ratings on this scale were significantly correlated with a measure of program evaluation utilization.

Need to Evaluate: Also from Weeks (1979), a sub-scale constructed in the same format. The two items were "Prior to this evaluation, did the decision makers directly
associated with this program see a need to change?" and
"Prior to this evaluation, did the decision makers directly
associated with this program see any need to change
specific policies that pertain to the agency as a whole?". Weeks (1979) also found ratings on this scale to be
significantly related to program evaluation utilization.
CHAPTER III
RESULTS

Extent and Type of Program Evaluation Utilization

In null form, the two hypotheses assessing the extent and type of program evaluation utilization found in the present study were:

H1: The probability of occurrence of program evaluation utilization will be zero.

H2: The probability of occurrence of the three types of utilization (conceptual, instrumental, and persuasive) will be equal.

In the current study, the extent of program evaluation utilization was measured using the following variables: 1) evaluator's report of the occurrence or non-occurrence of instrumental, conceptual and persuasive program evaluation utilization (Leviton and Hughes, 1981); 2) item response frequencies on Weeks' (1979) Program Evaluation Utilization Scale; 3) total number of items endorsed on Johnson's (1980) Program Evaluation Utilization Scale; and, 4) item endorsement means on the two utilization scales.

As indicated in Table 8, the 332 program evaluators surveyed in the present study who had conducted recent Canadian program evaluations and were aware of how the
Table 8

Extent of Recent Canadian Program Evaluation Utilization by Type

<table>
<thead>
<tr>
<th>TYPE OF UTILIZATION</th>
<th>RELATIVE PERCENT UTILIZED</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conceptual</td>
<td>65.3%</td>
</tr>
<tr>
<td>Instrumental</td>
<td>55.3%</td>
</tr>
<tr>
<td>Persuasive</td>
<td>40.8%</td>
</tr>
</tbody>
</table>

ABSOLUTE PERCENT

All Types          98.8%

Note. N = 332 recent Canadian program evaluations.

*A single evaluation could be reported as having been utilized in more than one manner so these relative percentages do not add to 100%. Indicating that one type of utilization had occurred did not preclude the subjects from indicating that either of the other two types had also occurred.*
results were utilized reported that program evaluation utilization of some type occurred in more than 98% of the evaluations on which they reported. Using this measure of utilization, the probability of utilization of some type was $p = .99$. When this probability is converted to a $z$ score for a sample of this size ($N = 332$), with $p$ and $q$ each set at $.5$ ($z = 17.8$), the likelihood of observing such a finding due to chance alone is less than $.001$.

As noted in Table 8, the occurrence or non-occurrence of the three types of utilization was not independent. The evaluators who participated in this study were free to indicate that either no utilization of any type has occurred or that utilization of only one, any two or all three types of utilization had occurred. Table 9 summarizes these data. No utilization of any type occurred in 1.2% of the sample. Utilization of a single type occurred in 47.0% of the sample. Utilization of two types occurred in 41.6% of the sample and utilization of all three types occurred in 10.2% of the sample.

Furthermore, as indicated in Tables 10 and 11, when the number of items endorsed on the Weeks (1979) scale and the item response frequencies on both the Weeks (1979) and Johnson (1980) scales were examined, utilization was observed in 91.3% and 97.6% of the 332 cases respectively. Therefore, the observed probabilities of the occurrence of utilization observed using these measures were $.91$ and $.98$, 

Table 9

Frequency of Different Utilization Types

<table>
<thead>
<tr>
<th>NUMBER OF TYPES</th>
<th>FREQUENCY</th>
<th>PERCENTAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>NONE</td>
<td>4</td>
<td>1.2%</td>
</tr>
<tr>
<td>ONE</td>
<td>156</td>
<td>47.0%</td>
</tr>
<tr>
<td>TWO*</td>
<td>138</td>
<td>41.6%</td>
</tr>
<tr>
<td>THREE</td>
<td>34</td>
<td>10.2%</td>
</tr>
</tbody>
</table>

Note. N = 332 recent Canadian program evaluations.

*When an evaluation was reported as being used as two types of utilization, it could be in one of these three pairs: instrumental and conceptual; instrumental and persuasive; or, conceptual and persuasive utilization.
Table 10

Weeks Utilization Scale: Item Response Frequencies

<table>
<thead>
<tr>
<th>ITEM NUMBER</th>
<th>PERCENT OF TOTAL ENDORSEMENTS</th>
</tr>
</thead>
<tbody>
<tr>
<td>1*</td>
<td>1.2%</td>
</tr>
<tr>
<td>2*</td>
<td>3.3%</td>
</tr>
<tr>
<td>3*</td>
<td>4.2%</td>
</tr>
<tr>
<td>4</td>
<td>7.5%</td>
</tr>
<tr>
<td>5</td>
<td>18.7%</td>
</tr>
<tr>
<td>6</td>
<td>3.9%</td>
</tr>
<tr>
<td>7</td>
<td>15.4%</td>
</tr>
<tr>
<td>8</td>
<td>12.0%</td>
</tr>
<tr>
<td>9</td>
<td>18.7%</td>
</tr>
<tr>
<td>10</td>
<td>15.1%</td>
</tr>
</tbody>
</table>

Note. N = 332 recent Canadian program evaluations.

*Endorsement of any of these three items indicates that no utilization of any type occurred. The total for all 3 items is therefore 8.7% non-utilization in a sample of 332 program evaluations.
<table>
<thead>
<tr>
<th>TOTAL NUMBER OF ITEMS ENDORSED</th>
<th>PERCENT OF TOTAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>0&lt;sup&gt;a&lt;/sup&gt;</td>
<td>2.4%</td>
</tr>
<tr>
<td>1</td>
<td>1.8%</td>
</tr>
<tr>
<td>2</td>
<td>5.1%</td>
</tr>
<tr>
<td>3</td>
<td>15.7%</td>
</tr>
<tr>
<td>4</td>
<td>23.2%</td>
</tr>
<tr>
<td>5</td>
<td>28.3%</td>
</tr>
<tr>
<td>6</td>
<td>18.4%</td>
</tr>
<tr>
<td>7</td>
<td>5.1%</td>
</tr>
</tbody>
</table>

Note. N = 332 recent Canadian program evaluations.

<sup>a</sup>0 items endorsed indicates that no utilization of any type occurred.
respectively. When these probabilities were converted to z scores, with p and q both set at .5, the likelihood of observing probabilities of these magnitudes in a sample of this size (N = 332) solely due to chance was less than .001 (z = 14.9 and 17.5 respectively).

Furthermore, as illustrated in Table 8, conceptual utilization was observed most frequently, occurring in 65.3% of the cases. Instrumental program evaluation utilization also occurred in more than half (55.3%) of the sample. Persuasive program evaluation utilization, whose existence had not been previously demonstrated empirically, was also reported to have occurred by 40.8% of this sample. A Chi Square analysis of these occurrence rates (Table 12) indicated that they differed significantly, $\chi^2 (2, N = 332) = 40.3, p < .001$. Pairwise Chi Square comparisons also yielded significant differences between conceptual and instrumental utilization, $\chi^2 (2, N = 332) = 4.74, p < .05$; instrumental and persuasive utilization, $\chi^2 (2, N = 332) = 14.4, p < .001$; and, conceptual and persuasive utilization, $\chi^2 (2, N = 332) = 39.6, p < .001$.

When the means of the two utilization scales were examined, it was observed for the Weeks (1979) scale that the mean number of items endorsed was 4.4 out of a possible total of 7. On the Johnson (1980) scale, the mean was 6.9 out a possible total of 10.
Table 12

Chi Square Analysis of Utilization Occurrence by Type

<table>
<thead>
<tr>
<th>UTILIZATION FREQUENCY</th>
<th>CONCEPTUAL</th>
<th>INSTRUMENTAL</th>
<th>PERSUASIVE</th>
</tr>
</thead>
<tbody>
<tr>
<td>UTILIZED</td>
<td>216 (65.3%)</td>
<td>184 (55.3%)</td>
<td>135 (40.8%)</td>
</tr>
<tr>
<td>NOT</td>
<td>116 (34.7%)</td>
<td>148 (44.7%)</td>
<td>197 (59.2%)</td>
</tr>
<tr>
<td>TOTAL</td>
<td>332</td>
<td>332</td>
<td>332</td>
</tr>
</tbody>
</table>

Note. N = 332 recent Canadian program evaluations.

Variables Associated with Occurrence of Program Evaluation Utilization Type

Hypothesis #3, in null form, stated that the evaluation process, evaluator and decision context variables observed in this study would be unrelated to the occurrence or non-occurrence of the three types of program evaluation utilization (instrumental, conceptual and persuasive). In order to test this hypothesis, the relationships between the occurrence of program evaluation utilization of these three types and the fifty-three evaluation process, evaluator and decision context variables were analyzed and tested for significance.

The three dichotomous indices of the occurrence/non-occurrence of the three types of utilization were used as the grouping variable in a series of crosstabulation procedures with nominal level and dichotomous evaluation process, evaluator and decision context variables (with significance levels tested using the Chi Square statistic). The results of these analyses are presented in Table 13.

The three dichotomous occurrence/non-occurrence indices were also used as the grouping variable in a series of SPSSPC MEANS procedures with interval level evaluation process, evaluator and decision context variables (with significance levels tested using the F test statistic). The
Table 13

Utilization Type Occurrence and Nominal-level and Dichotomous Evaluation Process, Evaluator and Decision Context Variables

<table>
<thead>
<tr>
<th>INSTRUMENTAL</th>
<th>CONCEPTUAL</th>
<th>PERSUASIVE</th>
</tr>
</thead>
<tbody>
<tr>
<td>A) EVALUATION PROCESS VARIABLES:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Presence of E's Recommendations***</td>
<td>Presence of E's Recommendations*</td>
<td></td>
</tr>
<tr>
<td>Timeliness***</td>
<td>Timeliness*</td>
<td></td>
</tr>
<tr>
<td>B) EVALUATOR VARIABLES:</td>
<td></td>
<td>E's Current Discipline*</td>
</tr>
<tr>
<td>E's Work Setting*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Practice Pattern</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Judged As Most</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Similar (Academic)*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C) DECISION CONTEXT VARIABLES:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Program Size*</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. Significant Chi Square statistics indicate a significant association between the noted variable and the occurrence of utilization of that type. Variables for which a non-significant relationship was observed were omitted from this table.

*p < .05. **p < .01. ***p < .001.
results of these analyses are presented in Table 14.

The chi square results reported in Table 13 and the F test results reported in Table 14 provide some support for the alternative form of Hypothesis #3, since a number of the nominal-level, dichotomous and interval-level variables were significantly related to the occurrence of the three different types of program evaluation utilization.

Variables Associated with Instrumental Utilization

Two nominal-level, one dichotomous and twelve interval-level variables out of fifty-three evaluation process, evaluator and decision context variables showed significant associations with the occurrence or non-occurrence of instrumental utilization.

Timeliness of presentation of the evaluation's findings was significantly associated with instrumental utilization, \( \chi^2 (3, N = 332) = 25.4, p < .001 \). Evaluation reports that were delivered on time were more likely to be utilized instrumentally than those which were delivered late or early. The presence or absence of specific action recommendations in the evaluation report was also significantly associated with instrumental utilization, \( \chi^2 (3, N = 332), p < .001 \). Evaluation reports which contained specific action recommendations were more likely to be used
Table 14

Occurrence of Utilization and Interval-level Evaluation

Process, Evaluator and Decision Context Variables

<table>
<thead>
<tr>
<th>Instrumental</th>
<th>Conceptual</th>
<th>Persuasive</th>
</tr>
</thead>
</table>

A) Evaluation Process Variables:

No. of Contact Hrs: Planning***

No. of Contact Hrs: Implementation**

No. of Contact Hrs: Dissemination**

Relevance of Findings***     Relevance of Findings***

Behavior Checklist***        Behavior Checklist**

B) Evaluator Variables:

E's Evaluation Experience**

E's Specialization in Evaluation***

E's Age*

Similarity to Academic Practice Pattern*
Table 14 (contd.)

<table>
<thead>
<tr>
<th>INSTRUMENTAL</th>
<th>CONCEPTUAL</th>
<th>PERSUASIVE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

C) DECISION CONTEXT VARIABLES:

Org'l. Constraints: Org'l. Cons.:  
Planning* Planning*  
Org'l. Cons.: Org'l. Cons.:  
Dissemination** Dissemination**  
Mgmt. Support*** Mgmt. Sup.***  
Need to Evaluate* Need to Evaluate*

Note. Significant F test statistics indicate a significant association between the noted variable and the occurrence of utilization of that type. Variables for which a non-significant relationship was observed were omitted from this table.

*p < .05. **p < .01. ***p < .001.
instrumentally than reports which did not include such recommendations. Finally, the significant result for program size, defined by the number of staff and volunteers in the program, which was also significantly related to instrumental utilization, $\chi^2 (2, N = 332) = 12.6, p < .03$, reflected the fact that instrumental utilization was more likely to occur in programs with more staff and volunteers.

Of the 12 interval-level variables, the scores on 5 evaluation process, 3 evaluator and 4 decision context variables were significantly associated with instrumental utilization. With regard to the 5 interval-level, evaluation process variables, evaluations were more likely to be utilized instrumentally when: the evaluator spent more contact hours with program decision makers during the planning ($M_{occurrence} = 24.5$ hours versus $M_{non-occurrence} = 11.4$ hours) with $f(1, 289) = 11.1, p < .001$; implementation ($M_{occ} = 59.2$ hours versus $M_{non} = 25.3$ hours) with $f(1, 288), p < .01$ and dissemination phases ($M_{occ} = 22.3$ hours versus $M_{non} = 10.8$ hours) of the evaluation with $f(1, 289) = 9.4, p < .01$; the evaluation findings were rated as more relevant to decision maker's needs ($M_{occ} = 2.5$ versus $M_{non} = 1.9$) by the evaluator, $f(1, 326) = 53.6, p < .001$; and, the evaluator performed more behaviors on a utilization-enhancing behavioral checklist ($M_{occ} = 12.2$ behaviors versus $M_{non} = 10.7$ behaviors), $f(1, 327) = 28.1, p < .001$. 
With regard to the 3 interval-level, evaluator variables, evaluations were more likely to be utilized instrumentally when: they were performed by evaluators with more evaluation experience, defined as evaluators who had performed significantly more past evaluations ($M_{occ} = 14.9$ past evaluations versus $M_{non} = 10.5$ past evaluations), $F(1, 325) = 7.0$, $p < .01$; they were performed by evaluators who were evaluation specialists, defined as evaluators who spent a significantly greater percentage of their total work time ($M_{occ} = 53.0\%$ versus $M_{non} = 37.9\%$) in conducting evaluations, $F(1, 325) = 17.2$, $p < .001$; and, when they were performed by older ($M_{occ} = 42.5$ years versus $M_{non} = 40.4$ years) evaluators, $F(1, 322) = 5.5$, $p < .02$.

With regard to the 4 interval-level, decision context variables, evaluations were more likely to be utilized instrumentally when: they were performed in decision contexts which evaluators rated as having fewer organizational constraints on evaluation planning ($M_{occ} = .98$ versus $M_{non} = 1.32$), $F(1, 328) = 4.9$, $p < .03$ and dissemination ($M_{occ} = 1.06$ versus $M_{non} = 1.58$), $F(1, 327) = 9.4$, $p < .01$; and, they were performed in decision contexts which evaluators rated as having more management support for the evaluation ($M_{occ} = 5.2$ versus $M_{non} = 3.8$), $F(1, 323) = 99.7$, $p < .001$ and more perceived need to evaluate ($M_{occ} = 3.5$ versus $M_{non} = 3.1$), $F(1, 323) = 6.0$, $p < .02$. 
Variables Associated with Conceptual Utilization

Three nominal-level, one dichotomous and four interval-level variables showed significant associations with the occurrence or non-occurrence of conceptual utilization. The evaluator’s primary work setting was observed to be significantly associated with conceptual utilization with $\chi^2 (6, N = 330) = 13.0, p < .05$, in that evaluators whose primary work setting was a university or a provincial government were more likely to report conceptual utilization than those whose work setting was the federal government or a private consulting firm.

With regard to evaluation practice pattern with $\chi^2 (3, N = 325) = 8.7, p < .04$, evaluators who judged themselves to be most similar to an academic pattern of evaluation practice were more likely to report conceptual utilization than those evaluators who judged themselves as most similar to one of the other three practice patterns. In contrast to the results for instrumental utilization, presence of specific recommendations decreased the likelihood of conceptual utilization with $\chi^2 (1, N = 331) = 4.1, p < .05$. In addition, and somewhat surprisingly, evaluations whose reports were not on time were more likely to be utilized conceptually with $\chi^2 (3, N = 329) = 10.1, p < .02$, than those in which the report was delivered on time.
Of the interval-level variables, scores on one evaluation process, one evaluator and two decision context variables were significantly associated with conceptual utilization. Evaluations which were utilized conceptually were rated as having produced more relevant findings ($M_{occ} = 2.5$ versus $M_{non} = 2.1$) with $t (1, 327) = 8.4$, $p < .02$ than those which were not utilized conceptually. In addition, evaluations which were utilized conceptually were performed in decision contexts which evaluators rated as having less management support ($M_{occ} = 4.4$ versus $M_{non} = 4.9$) for the evaluation, with $t (1, 323) = 12.7$, $p < .001$, and more organizational constraints on the planning ($M_{occ} = 1.26$ versus $M_{non} = 0.89$) of the evaluation with $t (1, 328) = 5.4$, $p < .02$ than those decision contexts in which evaluations were reported as not having been utilized conceptually. Finally, evaluations which were utilized conceptually were performed by evaluators who rated themselves significantly more similar to an academic evaluation practice style ($M_{occ} = 0.87$ versus $M_{non} = 0.63$) with $t (1, 325) = 3.9$, $p < .05$ rather than by evaluators who rated themselves as less similar to this evaluation practice style.
Variables Associated with Persuasive Utilization

Only one nominal-level and two interval-level variables showed significant associations with the occurrence or non-occurrence of persuasive utilization. One was an evaluation process variable and the other two were evaluator variables. Evaluations that were utilized persuasively were performed by evaluators who indicated that they had done significantly more of the behaviors listed on a behavioral checklist ($M_{occ} = 12.0$ behaviors versus $M_{non} = 11.3$ behaviors) with $f(1, 326) = 9.1, p< .01$ and, who were more experienced, defined as having conducted significantly more evaluations in the past ($M_{occ} = 16.0$ past evaluations versus $M_{non} = 10.8$ past evaluations) with $f(1, 325) = 9.1, p< .01$, than evaluators who did fewer of the behaviors on the checklist and had conducted fewer evaluations in the past. In addition, evaluations utilized persuasively were more likely to have been performed by evaluators who listed their current disciplinary affiliation as agriculture, business administration, organizational development and social work with $\chi^2 (14, N = 330) = 25.5, p< .03$, than those who listed other current disciplinary affiliations.
Patterns of Association by Type of Utilization

Hypothesis #4, in null form, stated that no one group of the three groups of independent variables (evaluation process, evaluator or decision context characteristics) would be more highly related to the dependent variables (occurrence of the three types of utilization) than any other group. The evaluation process, evaluator and decision context variables employed in the current study showed varying degrees of association with the three types of program evaluation utilization (Figure 1). Both instrumental and conceptual utilization were positively related to the following evaluation process variables: a rating made by the evaluator of how relevant the evaluation findings were for the decision makers' needs, evaluation reports that were completed on time and evaluation reports which contained recommendations.

By contrast, while instrumental utilization was negatively related to the decision context variable of extent of organizational constraints on the planning of the evaluation and positively related to the extent of management support for the evaluation, the reverse of these two relationships was found for conceptual utilizations. In addition, conceptual utilizations were more likely to be found in provincial government programs, while instrumental utilizations were more likely to be found in federal
Figure 1

Patterns of Association Between Utilization Type and Evaluation Process, Evaluator and Decision Context Variables

**INSTRUMENTAL UTILIZATION**
- Program Size
- No. of Contact Hrs.: Planning, Implementation and Dissemination
- E.'s Specialization in Evaluation
- Evaluator's Age
- Org'l Constraints on Dissemination
- Need to Evaluate

**CONCEPTUAL UTILIZATION**
- Evaluator's Primary Work Setting
- Single Practice Pattern (of four)
- Judged as Most Similar
- Rating of Similarity to the Academic Practice Style

**SHARED**
- Presence of Evaluator's Recommendations
- Timeliness
- Org'l Cons.: Planning Relevance
- Management Support

**PERSUASIVE UTILIZATION**
- Evaluator's Current Discipline

Evaluator's Experience
- Behavior Checklist
government program evaluations.

Both instrumental and persuasive utilizations were related to the evaluation process variable of higher scores on a behavioral checklist of evaluator behaviors purported to enhance utilization success (adapted from Brown & Braskamp, 1980). The evaluator variable which was positively related to both instrumental and persuasive utilization was the number of past evaluations that the evaluator had conducted, an index of the evaluator's level of experience.

Only instrumental utilizations were significantly related to the following variables: decision context -- programs of larger size, contexts with less constraints on the dissemination of findings, a higher need to evaluate; evaluator variables -- a greater percentage of the evaluator's time spent on conducting evaluations, slightly older evaluators; evaluation process variables -- number of contact hours with program decision makers during the planning and implementation of the evaluation.

Only conceptual utilizations were related to the following evaluator variables: the greater likelihood of occurrence with university-based program evaluators and those program evaluators who rated themselves as more similar to the academic pattern of evaluation practice.

Instances of persuasive program evaluation utilization showed only one unique significant relationship with the
evaluator variable of current disciplinary affiliation. Evaluators who listed their current discipline as either agriculture, business administration, organizational development or social work were more likely to report persuasive utilization.

In summary, fifteen of fifty-three independent variables were found to be related to instrumental utilization. Only five of these same fifteen variables were also significantly related to conceptual utilization and only two shared a relationship with the occurrence of persuasive utilization. In addition, the decision context variables of organizational constraints on the planning of the evaluation and extent of management support ratings made by evaluators were negatively related to instrumental utilization but positively related to conceptual utilization.

Eight of the fifty-three evaluation process, evaluator and decision context variables were significantly related to conceptual utilization with three being related to this type of utilization only.

Only three of the same fifty-three variables were found to be significantly related to instances of persuasive utilization. This finding and other nonsignificant trends in the data provide support for Leviton and Hughes' (1981) contention that this type of utilization is quite different from the other two.
Multivariate Analyses of Variance by Utilization Type

In order to further test Hypothesis #4, which predicted that one of the three groups of the independent variables, the characteristics of the evaluation process, would be more highly related to utilization than the other two groups of independent variables (the evaluator and decision context variables), a series of three Multivariate Analyses of Variance (MANOVAS) were performed. A subset of four variables were selected from each of three groups of independent variables (evaluation process, evaluator and decision context characteristics) for use in these analyses based upon three selection criteria. First, that they be interval-level variables. Second, that the distribution of scores on the variables closely approximate a normal distribution. Third, that the variables within each group not be too highly intercorrelated. Using these three selection criteria, twelve independent variables were chosen for inclusion in the three MANOVAS. These variables are listed by category in Table 18.

In order to control for the non-independence of occurrence of the three utilization types, only those cases in which a single type of utilization occurred were considered for use in these analyses (n = 156). Only conceptual utilization occurred in 83 of the cases with only instrumental utilization occurring in 64 of cases and
Table 15

INDEPENDENT VARIABLES BY CATEGORY USED IN MANOVAS

EVALUATION PROCESS VARIABLES:
Number of Contact Hours with Program Decision Makers:
Dissemination Phase
Percentage of Total Evaluation Time on Spent on Outcome
Score on Utilization-Enhancing Behavioral Checklist
Rating of the Relevance of the Evaluation's Findings

EVALUATOR VARIABLES:
Percent of Evaluator's Total Work Time Doing Evaluations
Evaluator's Age
Number of University Degrees Held by the Evaluator
Number of Past Evaluations Conducted by the Evaluator

DECISION CONTEXT VARIABLES:
Primary Program Decision Maker's Tenure
Rating of Organizational Constraints on Evaluation Planning
Rating of Need to Evaluate
Findings of the Current Evaluation
only persuasive utilization occurring in 9 cases. Due to the small number of cases of solely persuasive utilization, these cases were excluded from the three MANOVAS. The remaining 147 cases of "pure" instrumental and conceptual utilization were used in a series of three MANOVAS as independent variables (factors) with values of 0 or 1.

The results of the three one way MANOVAS revealed no significant findings for any of the three groups of dependent variables (selected evaluation process, evaluator and decision context characteristics). Therefore, Hypothesis #4 was again not supported by these additional analyses. These additional analyses again failed to support the alternative hypothesis that evaluation process characteristics would be more highly related to the extent of utilization than the other two groups of independent variables.

Other Hypotheses

The present study included three other posited hypotheses regarding the relationships between specific potential influencing variables and the evaluation utilization criteria. These hypotheses were derived from the results of studies of American program evaluation utilization and on the theoretical speculations of many
American and Canadian evaluators. An attempt was made to replicate these findings in the Canadian program evaluation context. The results of the tests of these three hypotheses follow.

**Hypothesis #5**

Hypothesis #5, based on Patton et al. (1978) and others, in its alternative form, predicted that evaluators who practised a stakeholder-oriented, participative style of program evaluation practice would achieve greater utilization than evaluators who practised other styles of evaluation research. To test this hypothesis, evaluators were divided on the basis of the one of four evaluation practice patterns (academic, effectiveness judge, internal change agent, stakeholder-oriented) that they judged as most similar to their own evaluation practice pattern. The utilization scores achieved by the four groups of evaluators on the two utilization scales and the proportion of occurrences of instrumental, conceptual and persuasive utilization were then compared using Chi squares and F tests. This hypothesis was not supported.

Only one significant difference in utilization between the four groups was observed and it was not the relationship predicted by Hypothesis #5. Based upon 322
cases (10 subjects did not respond to this item), the evaluators who indicated that their evaluation practice was most similar to the Stakeholder-Oriented Practice Pattern (Pattern C) actually reported significantly fewer instances of conceptual utilization with \( \chi^2 (3, N = 322) = 7.83, p < .05 \) and \( f (1,4) = 2.46, n = 322, p < .05 \) as most similar to a stakeholder-oriented with only 47% of their evaluations utilized conceptually whereas the other three groups of evaluators reported conceptual utilization in between 64% and 73% of their evaluations. No significant differences were found among the other three evaluation practice patterns. Evaluators who indicated that their practice style was most similar to the stakeholder style reported more instrumental utilization (61% versus 55%, 55% and 52% for the other three styles). This was not a significant difference, however. Finally, the four groups did not differ significantly on their scores on either the Weeks (1979) or Johnson (1980) utilization scales.

Hypothesis #6

A number of American program evaluators (e.g., Ciarlo, 1981; Stevenson, 1981) have contended that evaluators who are internal to the program being evaluated will achieve greater utilization (especially of the instrumental type)
than external evaluators. This contention was based on American case study data only. This prediction (Hypothesis #6) was tested by recoding the evaluator's organizational position data as either internal or external and then comparing the dependent variables for these two groups of evaluators. *No significant differences* on any of the five measures of utilization were found between these two groups when a chi square test of association was performed. As previously mentioned, the vast majority of the evaluators were external to the program that they evaluated (approximately 88%). This fact may have reduced the possibility of obtaining significant differences between the two groups.

**Hypothesis #7**

Patton et al. (1978), in discussing their findings with regard to twenty American federal program evaluations, concluded that identifying a key program decision maker to take responsibility for utilization would significantly increase the likelihood that the evaluation's findings would be utilized. This hypothesis was tested by dividing the study participants on the basis of their answer to the item: "I attempted to identify a key program decision maker who would take responsibility for utilization." Two
analyses were designed to test the significance of differences on the five utilization measures between those who answered "yes" to this item and those who answered "no." The statistical procedures employed were a Crosstabs procedure with a Chi Square test of significance and an Analysis of Variance procedure with an F test of significance.

This hypothesis received partial support from the data obtained from this sample of 332 recent Canadian program evaluation studies. These results are presented in Table 16. Evaluators who indicated that they had identified key program decision makers as potential users did not report significantly greater conceptual or persuasive utilization than those who did not report identification of key program decision makers and evaluation users.

However, those evaluators who identified key program decision makers and evaluation users did report a significantly greater extent of instrumental utilization (59.3% versus 39.4%) with $\chi^2 (1, N = 329) = 8.47, p< .01$ than did those evaluators who did not. In addition, these same evaluators obtained higher scores ($M_{ID} = 4.48$ versus $M_{noID} = 4.09$) on the Johnson Program Evaluation Utilization Scale with $f (1,331) = 7.795, p< .07$. No significant differences were observed between the two groups on the Weeks Program Evaluation Utilization Scale.
<table>
<thead>
<tr>
<th>Utilization and Identification of Key Decision Makers and Evaluation Users</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Utilization Criteria</strong></td>
</tr>
<tr>
<td>----------------------------</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Instrumental Utilization</td>
</tr>
<tr>
<td>Conceptual Utilization</td>
</tr>
<tr>
<td>Persuasive Utilization</td>
</tr>
<tr>
<td>Weeks Scale Means</td>
</tr>
<tr>
<td>Johnson Scale Means</td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

**Note.** N = 329 recent Canadian program evaluations.

*p < .01  **p < .07.
In summary, two hypotheses designed to test the replicability of some American evaluation utilization findings were not generalizable in the Canadian program evaluation context. These results support the contention made by a number of Canadian program evaluators that the Canadian program evaluation context differs in some key ways from that found in the United States. A third hypothesis based upon American research conducted by Patton et al. (1978) received only partial support from data obtained from the present Canadian sample.
CHAPTER IV
DISCUSSION

Extent of Program Evaluation Utilization in Canada

The current study represented the first major attempt to determine the extent of recent program evaluation utilization in Canada. The results suggest that Canadian program evaluations were utilized extensively during the mid to late 1980s. A variety of measures demonstrated that between 91 and 99 percent of this sample of more than 300 recent Canadian program evaluations were utilized in some manner. As noted in Table 9, 47% of the 332 program evaluations included in the present sample had been utilized in at least one way. An additional 42% of the evaluations had been utilized in two different ways, and 10% of the respondents claimed that their evaluation findings had been utilized instrumentally, conceptually and persuasively.

Some caution should be exercised in generalizing these findings to the entire population of Canadian program evaluations. A number of factors may have operated concurrently to inflate these figures to at least some extent. For instance, only 85% of the program evaluators who had conducted a recent Canadian program evaluation and responded to the survey were aware of how the results of
their evaluation had been utilized. In at least some cases, this lack of awareness could reflect a failure to utilize evaluation results. In addition, it is possible that the program evaluators who responded to this survey were selective in the evaluations on which they chose to report. Although respondents were instructed to report on the last Canadian program evaluation they had completed between January 1, 1985 and January 1, 1988, it is impossible to assess the degree to which this instruction was followed. It seems plausible that program evaluators responding to the present survey might have been more inclined to report on their successes rather than on their failures.

Types of Program Evaluation Utilization in Canada

The current study found evidence for the existence of the three types of program evaluation utilization discussed by Leviton and Hughes (1981). The fact that many evaluations were utilized instrumentally and conceptually is not surprising, as the existence of these two types of utilization have been widely demonstrated by other researchers. However, persuasive program evaluation utilization, the use of the findings of a program evaluation to defend or attack a program, was reported for approximately 40% of the evaluations included in the
present study. This finding is an important one because the existence of this type of evaluation utilization has been a matter of debate in the evaluation utilization literature, and reports of persuasive utilization in the present study provide one of the first empirical demonstrations of its existence.

Another interesting finding in the current study was the observation that conceptual program evaluation utilization occurred significantly more frequently than instrumental utilization (65% and 55%, respectively). Most of the literature on evaluation utilization predicts that instrumental utilization should have been most common. The finding in the present study confirms Hypothesis #2 and also supports the minority view championed by evaluators such as Patton and Alkin, who contended that conceptual utilization is the most common type of program evaluation utilization.

Variables Associated with Program Evaluation Utilization

It was hypothesized that measures of evaluation process, evaluator and decision context variables would be associated significantly with measures of Canadian program evaluation utilization. This hypothesis was confirmed. However, differential levels of confirmation were observed
with the three types of utilization.

In the case of instrumental utilization, significant associations were observed with fifteen of the independent variables, including seven evaluation process variables, three evaluator variables and five decision context variables. For conceptual utilization, significant associations were found with eight independent variables including three evaluation process, three evaluator and two decision context variables. Finally, significant associations were observed between persuasive utilization occurrences and three independent variables including one evaluation process and two evaluator variables.

A number of plausible explanations exist for the difference in number of significant associations for each type of utilization. Instrumental utilization may have been more reliably measured than conceptual or persuasive utilization in the current study, thereby increasing the likelihood of other significant associations. Instances of instrumental utilization might be more clearly identifiable than instances of the other two types of utilization for a number of reasons. For example, the emphasis placed upon instrumental utilization in the evaluation literature may have predisposed evaluators to look for it. In addition, the definition of this type of utilization was probably more familiar to evaluators than the definitions of conceptual and persuasive utilization prior to their
participation in the current study.

Instances of conceptual utilization may have been less easily identifiable and therefore more likely to be either missed or identified inappropriately, thereby reducing the reliability of this dependent variable. There is also some evidence that the occurrence of conceptual utilization may have been viewed as a kind of consolation prize by study participants. Evidence for this possibility is provided by a Chi Square analysis of the classification of evaluations as instrumental or conceptual utilization. When the evaluation was classified as instrumental utilization the likelihood that is was also classified as conceptual utilization was less than 38%; when instrumental utilization was not judged to have occurred, the likelihood of a conceptual utilization was more than 91%. This finding was highly significant, $\chi^2 (1, N = 332) = 7.8.3, p < .001$.

The concept of persuasive utilization is probably even less familiar to evaluators than the idea of conceptual utilization. Furthermore, the selection of independent variables was based upon evaluation literature which has dealt largely with instrumental and conceptual utilization; therefore the independent variables that may in fact be associated with persuasive utilization may not have been included in the current study.

In addition to the constraints on the generalizability of the results discussed thus far, the nature of the
statistical analyses should also be considered. In total, 66 Chi Square analyses were performed using the three types of utilization and 22 nominal-level and dichotomous independent variables. In addition, a total of 90 F tests were performed using the three dependent variables and 30 interval-level independent variables. The sheer number of these statistical comparisons guarantees a number of significant associations between independent and dependent variables by chance alone. At the .05 level of significance, it can be expected that between three and four significant Chi Squares should be observed as well as four or five significant F tests.

In the current study, nine significant Chi Squares were observed, with seven of the nine being significant at the .05 level. Therefore, somewhat less than half of these significant associations may in fact have been due to chance alone. Both the timeliness variable and the presence of the evaluator's recommendations in the final report were observed to be significant at the .001 level, so confidence can be placed in these findings. However, the other seven independent variables which were found to be significant were all significant at the .05 level only.

Analyses of variance were significant in 18 of the 90 analyses whereas chance alone could be expected to produce approximately 25% of these significant associations. However, five of these associations were significant at the
.001 level and an additional six were significant at the .01 level. Most if not all of these 11 associations are likely to be significant despite the large number of statistical comparisons. The independent variables in question are indicated in Tables 13 and 14. Of the remaining seven associations, which were significant at the .05 level only, most could have been produced by chance due to the large number of individual comparisons made.

The Relative Importance of the Three Categories of Independent Variables

The alternative form of Hypothesis #4 stated that the evaluation process variables would be more highly related to measures of program evaluation utilization than would either the evaluation or decision context variables. This hypothesis was not supported by the results of this study. As shown in Tables 13 and 14, 25 independent variables from all three categories were found to be related to both instrumental and conceptual utilization. Eleven of these were evaluation process variables, eight were evaluator variables and seven were decision context variables. In addition, only one evaluation process variable was found to be associated with persuasive utilization, whereas two evaluator variables were also found to be significantly
related to this dependent variable. Clearly, in contrast to the hypothesis proposed in the current study, all three categories of independent variables included important variables. Further study of this question which includes refinement of the operationalization of the independent variables is indicated.

In a further attempt to test Hypothesis #4, a selected group of the interval-level independent variables from the evaluation process, evaluator and decision context categories were included in three multivariate analyses of variance (MANOVAS). Using only those instances of "pure" instrumental and conceptual utilization, three separate MANOVAS were performed, one with each of the three groups of independent variables. None of these analyses yielded significant results. Therefore, the null form of Hypothesis #4 could not be rejected. That is to say, no single group of the three groups of independent variables was any more or less important in accounting for the variance in instrumental or conceptual utilization. All three categories of the independent variables (evaluation process, evaluator and decision context characteristics) showed important relationships with evaluation utilization in this study. Based upon the findings in the current study, no single category can be viewed as more important than any other. These findings suggest that utilization is a complex phenomenon that is influenced by a large number
of factors in an interactive fashion. Further research on these issues is indicated.

Other Tested Hypotheses

The present study tested three hypotheses with regard to the relationship between individual predictors and Canadian program evaluation utilization success that had been suggested by other evaluators working in the American evaluation context. It has been speculated that the American and Canadian program evaluation contexts differ substantially and that these differences might influence the potential replicability of the results of American research on program evaluation utilization. The results of this study lend credence to this speculation.

First, Patton et al. (1978), Alkin et al. (1979), Greene (1988) and others have reported that American evaluators who practised a stakeholder-oriented style of evaluation practice were more likely to achieve utilization success. This finding (Hypothesis #5) was not replicated in the current Canadian study. However, the failure to replicate this finding may be due in part to the fact that evaluator practice style may not have been operationally defined well in the current study. Evaluators who participated in the current study were asked to indicate
which of four paragraphs describing different evaluator practice styles was most similar to the style of evaluation practice that they had employed in the evaluation in question. These descriptions of evaluation practice style were based upon the survey findings of Shadish and Epstein (1987). More work on the refinement of these concepts may be warranted. In addition, only 39 of 332 evaluators who supplied information on this variable indicated that the stakeholder-oriented practice pattern was most similar to their own, thereby reducing the likelihood of observing significant findings.

Second, Ciarlo (1981) and others have speculated that evaluators who are internal to the program being evaluated are more likely to obtain utilization success when compared to external evaluators (Hypothesis #6). This hypothesis was also not supported in the Canadian context. However, it should be noted that only 12% of the total sample of evaluators were internal to the program, thereby reducing the likelihood of obtaining a significant difference.

Third, Patton et al. (1978) have argued that the key variable influencing evaluation utilization was whether or not the evaluator had identified a key program decision maker who would take responsibility for the utilization of the evaluation's findings (Hypothesis #7). This hypothesis received only partial support in the current Canadian study. Evaluators who identified a key decision maker who
would take responsibility for utilization obtained significantly more instrumental utilization but not significantly more conceptual or persuasive utilization. In addition, there was a non-significant trend for these same evaluators to have higher scores on the Johnson (1980) Utilization Scale which assesses the degree of instrumental utilization to at least some extent.

It must therefore be concluded that the Canadian evaluation context may in fact differ in some important ways from the American evaluation context thereby reducing the cross-border generalizability of research findings.

Relationships Among the Utilization Measures

It is important to examine the relationships among the five program evaluation utilization measures used in the current study for a number of reasons. First, evaluation theorists are not in total agreement with regard to which types of program evaluation utilization exist. Two major reviews of the program evaluation utilization literature (Leviton & Hughes, 1981; Cousins & Leithwood, 1986) have supported the distinction between instrumental and conceptual program evaluation utilization originally proposed by Rich (1977). However, only Leviton and Hughes (1981) have argued for the existence of persuasive program
evaluation utilization. Second, an examination of the data in the present study indicates that a pattern of 'interrelationships exists between the Weeks (1979) and Johnson (1980) utilization scales and the incidence of the three different types of utilization. Information on these interrelationships should be of use to other researchers who are considering using these two scales.

The findings of the current study lend support to Leviton and Hughes' (1981) hypothesis that three distinct types of utilization exist. First, when instrumental utilization was judged to have occurred, the probability that conceptual utilization would not occur was \( p = .89 \) and when conceptual utilization was present the probability that instrumental utilization had not occurred was \( p = .91, \chi^2 (1, N = 332) = 78.31, p < .001 \). Second, the occurrence or non-occurrence of persuasive utilization was unrelated to the occurrence or non-occurrence of either instrumental or conceptual utilization. It can therefore be concluded that three distinct types of utilization do exist.

With regard to the posited relationship between the three types of utilization and the two utilization scales, results are mixed. Weeks (1979) has contended that his utilization scale measures both instrumental and conceptual utilization. Scores on Weeks' scale were significantly positively related to the presence of instrumental utilization, \( r (332) = .57, p < .001 \). However, scores on the
Weeks scale were significantly negatively correlated with the presence of conceptual utilization, $r (332) = -0.23$, $p < .001$. Finally, scores on this scale were not correlated with the occurrence or non-occurrence of persuasive program evaluation utilization.

Johnson (1980) has also contended that his scale measures both instrumental and conceptual utilization. This contention also received partial support from the current study’s findings. As predicted, higher scores on the Johnson scale were significantly related to the occurrence of instrumental utilization, $r (332) = 0.44$, $p < .001$. However, in contrast to the Weeks scale, scores on Johnson’s (1980) scale were not related to the occurrence or non-occurrence of conceptual utilization but were positively significantly related to the occurrence of persuasive utilization, $r (332) = 0.23$, $p < .001$. This correlation is likely a result of the inclusion of a single item, item d, in the Johnson scale which describes a particular incidence of persuasive program evaluation utilization.

It should also be noted that scores on Weeks’ (1979) utilization scale were significantly positively correlated with Johnson’s (1980) scale scores, $r (332) = 0.52$, $p < .001$. This finding provides some evidence of the convergent validity of both scales as valid measures of program evaluation utilization. The two scales appear to share the
ability to measure instrumental program evaluation utilization to varying extents. Their use in future studies of instrumental utilization is therefore recommended.

Evidence for the Validity of the Three Types of Utilization

If the three types of utilization are valid concepts, they should be related to different patterns of the independent variables. Support for this view was provided by two main sources in the current study. First, the occurrence or non-occurrence of each of the three types of utilization was found to be related to different patterns of evaluation process, evaluator and decision context variables. These findings were summarized in Tables 13 and 14 and in Figure 1. For example, seven independent variables were associated solely with instrumental utilization, three were associated solely with conceptual utilization and one was associated only with persuasive utilization. Furthermore, a number of independent variables had directly opposing relationships with two types of utilization. For example, in comparing instrumental to conceptual utilization, instrumental utilization was more likely to have occurred when management support for the evaluation was high and when the organizational constraints on the planning of the evaluation were low. In direct
contrast, conceptual program evaluation utilization was more likely to have occurred when management support had been low and organizational constraints on planning had been high.

Second, as noted on page 139, direct comparisons of the occurrence rates of the three types of utilization through a series of Chi Square analyses showed that the presence of instrumental utilization was significantly associated with the absence of conceptual utilization. In addition, there were no significant relationships observed between the occurrence of persuasive utilization and either of the other two types of utilization.

In summary, evidence from a variety of sources lends support to the view that these three types of utilization are related to different patterns of the independent variables. This evidence also supports Leviton and Hughes' (1981) contention that three types of evaluation utilization exist.

Implications for Evaluation Practice

The results of this exploratory study have implications for the practice of Canadian program evaluators. For example, evaluators who seek to have their results utilized instrumentally should ensure that members
of the program management support their evaluation activities. In addition, evaluators should take the time early in the evaluation process to identify who will use the evaluation results and how they will be utilized. Evaluators must also ensure that they produce results that are delivered on time. The checklist of evaluator behaviors designed to enhance utilization adapted from Brown and Braskamp (1980) provides some valuable guidelines regarding which behaviors evaluators should perform if they hope to increase the likelihood that their results will be utilized. In addition, evaluators who wish to enhance the probability of achieving conceptual utilization of their findings should spend more on the dissemination phase of the evaluation and should use less rigorous research designs. Finally, evaluators who wish to enhance the likelihood that their results will be utilized persuasively should also spend more of their time on the dissemination of their findings.

Summary and Conclusions

Canadian program evaluations are utilized extensively. They are utilized in a variety of ways and each of the different ways in which they are utilized is related to different patterns of evaluation process, evaluator and
decision context variables.

The current study was predominantly exploratory and descriptive in nature but does suggest that Canadian program evaluation utilization is a topic worthy of further empirical investigation. Future studies of program evaluation utilization in Canada might profitably focus on the views of program decision makers and program line staff and not just evaluators as in the current study. It might also be of use to compare and contrast the differing viewpoints of evaluators and program stakeholders.

It is hoped that the current study will serve to inspire others to investigate the issue of program evaluation utilization in Canada.
REFERENCES


145


Washington, DC: The Urban Institute.


Appendix A
Personalized Cover Letter (French Version)

Michael P. Shea
Department of Psychology
University of Windsor
Windsor, Ontario
N9B 3P4

Le 20 septembre, 1988

(Subject’s Name & Address)

(Subject’s Name):

Je fais actuellement un sondage sur l’utilisation de l’évaluation de programme au Canada. Je suis un candidat au doctorat en psychologie appliquée à l’University of Windsor, travaillant sous la direction du Dr. Shelagh Towson. Depuis que j’ai mené ma première évaluation de programme il y a dix ans, je me suis intéressé aux facteurs qui influencent l’utilisation de l’évaluation de programmes au Canada. Ce sondage, qui est le sujet de ma these de doctorat, est une enquête sur ces facteurs, et je crois que les données que j’espère obtenir seront d’intérêt pour tous les évaluateurs canadiens. Puis-je vous demander de participer en remplissant le formulaire ci-joint? Vous trouverez ci-jointe une enveloppe adressée et timbrée.

Votre nom a été tiré du recent directoire de (Subject’s Evaluation Organization). En tant qu’évaluateur, vous êtes tout spécialement qualifié pour fournir des renseignements sur ce sujet essentiel.

Afin d’obtenir un échantillonnage complet, il est très important que vous remplissiez et retourniez tous les formulaires ci-joint.

Soyez assuré que vos réponses sont confidentiels. Votre identité, ainsi que celle de votre organisation resteront anonymes. Seulement les données globales de l’échantillon seront rapportées. Le sondage porte un numéro d’identification uniquement pour que votre nom soit reçu de la liste d’envoi quand il est retourné. Cette page sera détachée du sondage aussi tot qu’il est recu.

Si vous desirez recevoir un sommaire des résultats, veuillez me l’indiquer sur la page ou se trouve le formulaire d’acceptation.

Michael P. Shea
Appendix A
Personalized Cover Letter (English Version)

Michael P. Shea, M.A.
Department of Psychology
University of Windsor
401 Sunset Avenue
Windsor, Ontario
N9B 3P4

September 20, 1983

(Subject’s Name & Address)

Dear (Subject’s Name):

I am a Ph.D. student in Applied Social Psychology at the University of Windsor being supervised by Dr. Shelagh Towson. Ever since I conducted my first program evaluation research study more than ten years ago, I have had questions about what factors influence program evaluation utilization in Canada. I believe that information on this issue is of import to other Canadian program evaluators.

As a member of the (Subject’s Evaluation Organization), you are uniquely qualified to supply information on this vital topic. I am contacting all members of this organization. Your name was drawn from their membership list. To truly represent the population of Canadian program evaluators it is extremely important that each survey be completed and returned.

You may be assured of complete confidentiality. Your name and your organization will remain anonymous. The survey has an identification number for mailing purposes only. This is so I may check your name off the mailing list when the survey is returned. You may receive a summary of the results by indicating this preference on the consent form page of this survey which will removed from the survey upon receipt.

I or my dissertation supervisor, Dr. Shelagh Towson, would be most happy to answer any questions you might have. Please write or call. The telephone number is 519-253-4232, local 2250.
Thank you for your assistance.

Sincerely,

Michael P. Shea
APPENDIX B

CANADIAN PROGRAM EVALUATION UTILIZATION SURVEY

PSYCHOLOGY DEPARTMENT

CANADIAN PROGRAM EVALUATION UTILIZATION STUDY

164
THIS 20-PAGE SURVEY TAKES APPROXIMATELY 20 MINUTES TO COMPLETE. If you wish to participate in this study, please read and sign the Informed Consent Form on the next page, complete the survey, and mail it in the postage-paid self-addressed envelope provided. If you do not wish to participate, please return the survey so you will not be troubled with further mailings.

WHO AM I?

My name is Michael Shea. I am a Doctoral Candidate in Applied Social Psychology at the University of Windsor under the supervision of Dr. Shelagh Towson. I am a student member of the Canadian Evaluation Society (CES) and the Evaluation Section of the Canadian Psychological Association (CPA) and have been conducting program evaluation research since 1978.

WHAT IS THIS SURVEY FOR?

This survey is designed to obtain information about factors related to program evaluation utilization in Canada. Evaluation utilization is an important issue for program evaluators, given reports of high levels of non-utilization and the scarcity of resources for program evaluation. Unfortunately, we have little data on evaluation utilization and the data that do exist are based primarily on the experiences of American evaluators. The Canadian evaluation context differs in major ways from that in the U.S.; it is therefore vital that we collect some Canadian data. Your participation in this study will make an important contribution.

WHAT WILL BE DONE WITH THE SURVEYS?

The surveys and data will be stored in a secure place. Only aggregate results will be reported and disseminated. The information provided in this survey will provide the basis for my doctoral dissertation and for subsequent papers, presentations and publications. I will send a summary of the results of this survey to interested participants. No one will be informed if you choose not to respond. ALL OF THE INFORMATION THAT YOU SUPPLY IS CONFIDENTIAL.
IDENTIFICATION NUMBER: _____

THIS IDENTIFICATION NUMBER WILL ENSURE THAT AFTER YOU REPLY YOU WILL NOT BE CONTACTED ON THE FOLLOW-UP MAIL-OUTS. ASIDE FROM YOUR NAME AND ADDRESS, I HAVE NO INFORMATION ABOUT YOU OTHER THAN WHAT YOU PROVIDE ON THIS QUESTIONNAIRE. IF YOU HAVE ANY QUESTIONS OR CONCERNS, PLEASE CONTACT MICHAEL SHEA OR DR. SHELagh TOWSON AT 519-253-4232, EXTENSION 2250.

INFORMED CONSENT FORM

I, the undersigned understand that the purpose of this research is to obtain information on factors related to program evaluation utilization in Canada.

I understand that the information collected from me will be used only as part of a larger collection of information provided by other equally anonymous individuals and reported in group numerical or statistical form only. Thus, confidentiality will be safeguarded.

I agree to voluntarily participate in this study by completing the attached survey and returning it to the investigator no later than three weeks after receiving it.

I understand that this survey is a research undertaking being supervised through the Department of Psychology at the University of Windsor.

Date: _______________ Signature: _______________
(Print) Name: _______________

Thank you

WOULD YOU LIKE TO RECEIVE A SUMMARY OF THE RESULTS OF THIS STUDY? IF YES, PLEASE SUPPLY YOUR RETURN ADDRESS IN THE SPACE BELOW.

________________________________________
________________________________________

** Please note that this sheet will be detached from the survey upon receipt by the investigator.
PRELIMINARY QUESTIONS:

1) Did you participate as an evaluator in a Canadian program evaluation project completed between January 1, 1985 and January 1, 1988? A "program evaluation project" is defined here as an organized application of social scientific methodology to a determination of the effectiveness or efficiency of a program.

___ NO (IF NO, PLEASE "Stop" and mail this survey in the return envelope.)

___ YES (IF YES, PLEASE complete the remainder of the survey.)

2) What role did you play in the conduct of this evaluation?

___ Principal Investigator
___ Evaluation Project Director
___ Research Assistant
___ Other (specify): _______________________

IF YOU HAVE DETAILED KNOWLEDGE OF THE PROGRAM EVALUATION PROJECT AND THE PROGRAM THAT WAS EVALUATED PLEASE CONTINUE. IF YOU DO NOT HAVE THIS DETAILED KNOWLEDGE PLEASE STOP NOW AND MAIL THIS SURVEY.

3) Are you aware of how the results of your last evaluation (completed between January 1, 1985 and January 1, 1988) were utilized?

___ NO (IF NO, stop now and mail this survey).

___ YES (IF YES, Please continue with the survey).
PLEASE ANSWER ALL OF THE FOLLOWING QUESTIONS IN TERMS OF THE LAST EVALUATION PROJECT THAT YOU COMPLETED BETWEEN JANUARY 1, 1985 AND JANUARY 1, 1986.

1) Johnson (1980) Evaluation Utilization Scale:

Permission not given by Sage Publications to reprint this scale here. See references.
2) Weeks (1979) Utilization Scale:

Permission not given by Sage Publications to reprint this scale here. See references.
3) In your judgement, which of the following types of utilization occurred as a consequence of the evaluation you conducted (YOU MAY CHOOSE MORE THAN ONE ANSWER):

___ 0 No evaluation utilization occurred.

___ 1 Direct, instrumental uses of the findings in making program or policy decisions. Decision outcomes would have been different in the absence of the evaluation findings.

___ 2 Indirect, conceptual uses of the findings in discussing possible program changes or in planning changes in other programs that were similar to the program that was evaluated.

___ 3 Persuasive uses of the evaluation findings to defend or attack the program that was evaluated.

___ 4 Other type of utilization occurred (specify):

__________________________________________________
THE FOLLOWING QUESTIONS PERTAIN TO THE CHARACTERISTICS OF THE PROGRAM THAT YOU EVALUATED.

1) What type of program did you evaluate?
   ___ a. Federal government program
   ___ b. Provincial government program
   ___ c. Non-profit society program (governed by a single person)
   ___ d. Non-profit society program (governed by a administrative board)
   ___ e. For-profit company program (specify) ______
   ___ f. Other (specify) ________________________

2) How long had the program that you evaluated been in operation at the time of this evaluation?
   ___ a. Less than 1 year
   ___ b. 1-2 years
   ___ c. 2-3 years
   ___ d. More than 4 years
   ___ e. Don’t know

3) What is the total number of staff (including administrators, professional line staff, support staff, and volunteers) in the program that you evaluated?
   ___ a. Less than 5
   ___ b. 5-10
   ___ c. 10-20
   ___ d. 20-50
   ___ e. 50-100
   ___ f. More than 100
   ___ g. Don’t know

4) Where are most decisions made for the program that you evaluated?
   ___ a. Most decisions that affect program operations are made at the local program level
   ___ b. Most decisions that affect program operations are made by the program’s local, parent body
   ___ c. Most decisions that affect program operations are made by program’s non-local, parent body
   ___ d. Other (specify) ________________________
5) How long had the primary decision maker (i.e., program director or manager) of the program that you evaluated held his/her position at the time of this evaluation?
   ______ (Years)

6) How many contact hours did you have with the program's primary decision maker(s) during various phases of this evaluation:
   a) planning  ______ (Hours)
   b) implementation  ______ (Hours)
   c) post results period (feedback and discussion of findings)  ______ (Hours)

7) Who initiated this evaluation?
   ___ a. Primarily personnel involved with the program being evaluated
   ___ b. The program's parent organization
   ___ c. The program's funder or other overseeing body
   ___ d. Other (Specify): ____________________________

8) Who funded this evaluation?
   ___ a. Primarily personnel involved with the program being evaluated
   ___ b. The program's parent organization
   ___ c. The program's funder or other overseeing body
   ___ d. Other (Specify): ____________________________

9) How many previous evaluations of this program have been carried out (not counting the current evaluation)?
   ___ 0. No previous evaluations of this program
   ___ 1. One previous evaluation
   ___ 2. Two previous evaluations
   ___ 3. Other (specify) ________
   ___ 4. Don't know
10) If previous evaluations of this program have been completed, what was the general conclusion reached concerning the effectiveness of this program? Effectiveness was viewed as being:

__ a. Not at all effective
__ b. Minimally effective
__ c. Somewhat effective
__ d. Predominantly effective
__ e. Totally effective
__ f. Don't know
__ g. Not Applicable

11) How many program decision makers did you have to encourage or convince to use this evaluation's findings?

___

THE FOLLOWING QUESTIONS ADDRESS THE DEGREE TO WHICH ORGANIZATIONAL CONSTRAINTS IN THE ORGANIZATION IN WHICH YOU CONDUCTED THIS EVALUATION AFFECTED YOUR ABILITY TO CONDUCT THE EVALUATION. ORGANIZATIONAL CONSTRAINTS INCLUDE STAFF RESISTANCE, RESTRICTED ACCESS TO NEEDED INFORMATION OR DATA SOURCES, AND EVALUATION BUDGETARY CONSTRAINTS.

12) Did you experience organizational constraints during various phases of this program evaluation project?

<table>
<thead>
<tr>
<th></th>
<th>No</th>
<th>Yes</th>
<th>Minimal</th>
<th>Moderate</th>
<th>Extreme</th>
</tr>
</thead>
<tbody>
<tr>
<td>Planning</td>
<td></td>
<td></td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Implementation</td>
<td></td>
<td></td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Dissemination</td>
<td></td>
<td></td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
</tbody>
</table>
13) Did your evaluation report include specific recommendations for program change based upon the evaluation findings?

___ 0. No
___ 1. Yes

14) Did information from extraneous sources, not directly related to the program, compete with this evaluation's findings in the decision making process?

___ 0. No
___ 1. Yes
___ 2. Don't know

15a) Did you feel that most decision makers in this program and its parent body supported this evaluation of their program?

___ a. Yes, definitely
___ b. Yes, somewhat
___ c. No, not really
___ d. No, not at all
___ e. Don't know

b) Do you think that most of the key decision makers in this program and its parent body will use this evaluation's results as a guide for making improvements?

___ a. Yes, definitely
___ b. Yes, somewhat
___ c. No, not really
___ d. No, not at all
___ e. Don't know

16a) Prior to this evaluation, did the decision makers directly associated with this program see a need to change?

___ a. Yes, definitely
___ b. Yes, somewhat
___ c. No, not really
___ d. No, not at all
___ e. Don't know
b) Prior to this evaluation, did the decision makers directly associated with this program see any need to change specific policies that pertain to the agency as a whole?

___ a. Yes, definitely
___ b. Yes, somewhat
___ c. No, not really
___ d. No, not at all
___ e. Don't know

THE FOLLOWING ITEMS PERTAIN TO YOU AS AN EVALUATOR.

1) What percentage of your total work time do you spend in planning and conducting program evaluations?

_____ (%) 

2) What percentage of the total time that you spend in conducting program evaluations do you spend in each of the following settings?

___ % a. University
___ % b. Research or evaluation staff of an organization
___ % c. Provincial government
___ % d. Federal government
___ % e. Private consulting firm
___ % f. Independent private practice
___ % g. Other (Specify) ______________________

3a) Are you a member of the staff of the program that you evaluated?

___ 1. Yes
___ 2. No
b) IF YES, what is your position within the program?

__ 1. Member of the program's line staff, temporarily assigned to evaluation duties for this project only
__ 2. Member of the program's line staff with a permanent, part-time evaluation role
__ 3. Member of the program's permanent, full-time research or evaluation staff
__ 4. Other (specify) _______________________

c) IF NO, what is your position outside the program?

__ 1. Member of an evaluation unit directed by the program's parent organization
__ 2. External consultant hired directly by the program being evaluated
__ 3. External consultant hired by the program's parent organization
__ 4. Other (specify): _______________________

4-5) Please list the highest degree you hold and the associated discipline.

Degree _______________________
Discipline _______________________

6) What would you describe as your current primary disciplinary affiliation. Indicate ONE CHOICE ONLY.

__ 01 Agriculture
__ 02 Anthropology
__ 03 Business
__ 04 Economics
__ 05 Education
__ 06 Evaluation
__ 07 History
__ 08 Law
__ 09 Organizational Development
__ 10 Philosophy
__ 11 Political Science
__ 12 Psychology
__ 13 Public Administration
__ 14 Research & Statistical Methods
__ 15 Social Work
__ 16 Sociology
__ 17 Other (Specify) _______________________

7) Year of your birth: 19

8a) What is your gender?
___ 1. Female
___ 2. Male

b) What effect has your gender had on your evaluation work?

\[ \begin{array}{ccccc}
-2 & -1 & 0 & +1 & +2 \\
\text{Very Negative} & \text{Somewhat Negative} & \text{No Effect} & \text{Somewhat Positive} & \text{Very Positive} \\
\end{array} \]

9) How many program evaluations have you participated in as an evaluator (including the evaluation in question)?

\[(N) = \underline{\text{Program Evaluations}}\]

10) What percentage of your total evaluator effort in this evaluation was spent on each of the following activities? If you didn't participate in a phase please indicate this with the notation "N/A".

___ % a. Evaluability Assessment: determining if the evaluation was feasible, needed, and/or wanted.

___ % b. Process (formative) Evaluation: clarification of goals and objectives, monitoring program implementation and program processes.

___ % c. Outcome (summative) Evaluation: activities required to make a judgement as to the effectiveness of the program.

___ % d. Dissemination of evaluation findings.

___ % e. Other (specify) ____________________

11) PATTERN A: The evaluation was initiated to satisfy scientific interests and to serve long-term social theory and social problem solving purposes. The evaluation questions and outcome criteria were developed from the relevant literature or from considering the nature of the program itself. The evaluation used primarily traditional quantitative evaluation methods. Facilitation of utilization was attempted by making the evaluation results available in a public forum such as a scholarly journal, a conference presentation, or in the media.

<table>
<thead>
<tr>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not at all Similar to How I Conducted This Evaluation.</td>
<td>Somewhat Similar to How I Conducted This Evaluation.</td>
<td>Totally Similar to How I Conducted This Evaluation.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

12) PATTERN B: The evaluator was paid by the client to do the evaluation, in the hope that the evaluation would provide useful information to improve the program. The evaluator’s role was viewed as that of a local change agent or program team member. The evaluation questions and performance criteria for the program were formulated on the basis of stakeholders’ information needs. Methods used included participant and non-participant observation or interviews with stakeholders. Attempts to foster utilization involved extensive contact with the potential users throughout the course of the evaluation.

<table>
<thead>
<tr>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
</table>
13) **Pattern C**: The evaluator decided to evaluate after considering the potential benefits and changes that might result from the evaluation in the light of costs. This evaluator saw his or her role as a servant of program constituents but not as a local program team member or change agent. Evaluation questions and issues were formulated on the basis of pending decisions and legislation. No particular preference for any one set of methods was expressed. Utilization success was defined very strictly as direct, instrumental effects on decisions.

14) **Pattern D**: The evaluator saw the purpose of the evaluation as being to judge program effectiveness and worth. The evaluator's role is viewed as that of a methodological expert, educator of evaluation clients, and judge of program value. Both program monitoring and traditional quantitative methods were used. Attempts were made to facilitate utilization by providing both written and oral reports of final results complete with action recommendations.

15) Which of the preceding four patterns of evaluation practice style was most similar to how you conducted the evaluation?  (CHOOSE ONE ONLY)

- [ ] Pattern A
- [ ] Pattern B
- [ ] Pattern C
- [ ] Pattern D
16) Please consider the program evaluation writings that have shaped your evaluation practice. Please list up to three of these writings and rate the amount of influence that each had on your evaluation practice in this evaluation on a scale of 0 to 10 (0 = no influence, 5 = some influence, 10 = great influence):

Influence #1: ________________________________


RATING ____

Influence #2: ________________________________


RATING ____

Influence #3: ________________________________


RATING ____

17) FOLLOWING ARE A LIST OF POSSIBLE EVALUATOR BEHAVIORS. PLEASE INDICATE WHETHER OR NOT YOU PERFORMED EACH BEHAVIOR DURING THE COURSE OF THE EVALUATION.

a) I determined the extent of the program administrator's personal commitment to doing the evaluation project.

YES  NO

b) I determined my share of the responsibility for utilization.

YES  NO
c) I attempted to identify a key program decision-maker who would take responsibility for utilization.

YES      NO

d) I performed an evaluability assessment (that is, I determined if an evaluation was feasible and wanted before I decided to proceed with the evaluation).

YES      NO

e) I identified the decision makers and potential users of the evaluation information within and outside the organization.

YES      NO

f) I determined which specific decisions or policies might be influenced by the results of the evaluation.

YES      NO

g) I made sure that there was a clear understanding of the evaluation's role (the gathering of formative/process vs summative/outcome information).

YES      NO

h) I determined my freedom to provide evaluative information to program stakeholders.

YES      NO

i) I involved key program personnel in determining the purposes, issues, and general evaluation strategies.

YES      NO

j) I made sure that the data collection procedures were understandable and relevant to program stakeholders.

YES      NO
k) I had both informal and formal meetings with key program personnel.

YES  NO

l) I adapted the evaluation plan to meet the changing information needs of program stakeholders.

YES  NO

m) I shared rough drafts and recommendations with key program persons before writing the final report.

YES  NO

n) I attempted to suit the language and style of my reports and presentations to their intended audiences.

YES  NO

o) I did follow-up with evaluation clients to see if they had utilized the evaluation’s findings.

YES  NO

p) If the program personnel had not utilized the evaluation’s findings by follow-up, I attempted to change this with a further intervention.

YES  NO

PLEASE ANSWER EACH OF THE FOLLOWING QUESTIONS ABOUT THE CHARACTERISTICS OF THE EVALUATION PROCESS.

1) In your opinion, how relevant were the evaluation’s findings for the decision makers’ needs?

0. Not Relevant
1. Somewhat Relevant
2. Mostly Relevant
3. Totally Relevant
4. Don’t Know
2) What methodological strategy or strategies were used in this evaluation?

3) Based upon the current evaluation's findings, what was the overall conclusion about the program's effectiveness?

   a. Not at all effective
   b. Minimally effective
   c. Somewhat Effective
   d. Predominantly Effective
   e. Totally Effective

4) In your opinion, how timely was the presentation of the evaluation's results for organizational decision making needs?

   a. Too late to be utilized
   b. Somewhat late but still utilized
   c. On time
   d. Premature (too early to be utilized)
   e. Other (specify) ____________________________
   f. Don't Know

5) When was the evaluation in question completed?

   ___________________ (Month, Year)

6) Did you give a written final evaluation report to the primary decision maker in this program?

   No
   Yes
7) What was the total budget for this evaluation?

$ __________

ADDITIONAL COMMENTS:

********************************************************************

THIS IS THE END OF THE QUESTIONNAIRE. ONCE AGAIN, THANK YOU
VERY MUCH FOR YOUR COOPERATION, TIME, AND INFORMATION.
Appendix C

BILINGUAL REMINDER CARD

DEAR/CHÈR/CHÈRE COLLEAGUE/COLLABO:

APPROXIMATELY ONE WEEK AGO YOU SHOULDN'T HAVE RECEIVED A SURVEY ABOUT CANADIAN PROGRAM EVALUATION UTILIZATION. IF YOU HAVE ALREADY RETURNED THE SURVEY, THANK YOU VERY MUCH. IF YOU HAVE NOT YET RETURNED THE SURVEY PLEASE DO SO AS SOON AS POSSIBLE. THANK YOU VERY MUCH FOR YOUR TIME AND VALUABLE INFORMATION.

IL Y À UNE SEMAINE A PEU PRÈS, JE VOUS AI ENVOYÉ UN SONDAGE SUR L'UTILISATION DE L'ÉVALUATION DE PROGRAMMES AU CANADA. JE VOUS REMERCI SI VOUS AVEZ DÉJÀ RETOURNÉ LE SONDAGE. SI VOUS N'AVEZ PAS ENCORE RETOURNÉ LE SONDAGE, POURRAIS-JE VOUS DEMANDER DE LE FAIRE AUSSITÔT QUE POSSIBLE? JE TIENS À VOUS REMERCIER POUR VOTRE TEMPS ET POUR LES RENSEIGNEMENTS QUE VOUS FOURNISSEZ.

Michael P. Shea
Psychology Department
University of Windsor
Appendix D
Personalized Reminder Letter (French Version)

Michael P. Shea
Department of Psychology
University of Windsor
401 Sunset Avenue
Windsor, Ontario
N9B 3P4

Le 25 octobre, 1988

(Name & Address)

Madame/Monsieur,


A present, les evaluateurs de programme canadiens ne disposent pas de donnees sur l'utilisation de l'evaluation de programme. Il manque aussi des reseignements sur la gamme des facteurs qui influencent l'utilisation de l'evaluation de programme. Etant donne votre interet en evaluation, je crois que vous trouverez des reseignements sur ce sujet interessants ainsi qu'utiles. De plus, ces reseignements pourraient assister la croissance de l'evaluation de programme au Canada.

Afin d'obtenir un echantillonage complet, il est tres important que tous les formulaires du sondage soient remplis et retournes.

Si vous desirez recevoir un sommaire des resultats, veuillez me l'indiquer sur la page du formulaire d'acceptation.

Si vous avez des questions, n'hesitez pas a communiquer, avec moi, ou avec le Dr. Shelagh Towson; soit par ecrit, soit par telephone, au 519 253 4232, local 2250.

Je tiens a vous remercier d'avance pour votre participation. Veuillez agreer, madame/monsieur, l'expression de mes sentiments distingués.

Michael P. Shea
Appendix D
Personalized Reminder Letter (English Version)

Michael P. Shea
Department of Psychology
University of Windsor
401 Sunset Avenue
Windsor, Ontario
N9B 3P4

October 25, 1988

(Name & Address)

Dear Colleague:

Approximately four weeks ago you should have received a survey designed to obtain information about Canadian program evaluation utilization success. If you have already returned the survey, thank you very much. If not, please complete it now. Enclosed please find a second copy of this survey along with a postage-paid return envelope.

At present, Canadian program evaluators have no systematic information on how much or what types of program evaluation utilization are occurring. Nor do we have any information on which factors influence or determine those instances of utilization that do occur. As someone with an interest in evaluation, you will likely find that information on these issues will be of direct interest and use. In addition, information on these important issues may aid in the advancement of the discipline of program evaluation in Canada.

To truly represent the population of Canadian program evaluators it is extremely important that you respond.

You may receive a summary of the results by indicating this preference on the consent form page of the survey which will be removed immediately upon receipt by the investigator.

I or my dissertation supervisor, Dr. Shelagh Towson, would be most happy to answer any questions you might have. Please write or call. The telephone number is 519-253-4232, local 2250. Thank you for your assistance.
Sincerely,

Michael P. Shea
Appendix E
Letter to Department Heads (French Version)

Michael P. Shea, M.A.
Department of Psychology
University of Windsor
Windsor, Ontario
N9B 3P4

Le 20 septembre, 1988

(Department Head’s Name & Address)

(Name),

Je fais actuellement un sondage sur l’utilisation de l’évaluation de programme au Canada. Je suis un candidat au doctorat en psychologie sociale appliquée à l’University of Windsor, travaillant sous la direction du Dr. Shelagh Towson. Ce sondage est le sujet de ma these de doctorat. Puis-je demander votre collaboration en distribuant des formulaires du sondage aux membres de votre faculté et à vos étudiants au niveau supérieur qui sont actifs en evaluation de programme?

Dans le but d’obtenir le plus grand échantillon possible d’évaluateurs canadiens, je communique avec les chefs de tous les départements d’universités canadiens qui offrent un enseignement supérieur en criminologie, enseignement, administration d’enseignement, administration publique, psychologie, service social, et en sociologie.

Si vous desirez des copies additionnelles du sondage ou de plus amples renseignements, n’hésitez pas a me communiquer, soit par écrit, soit par telephone, au 519 253 4232, local 2219.

Je tiens à vous remercier d’avance pour votre concours.

Veuillez agréer, (Madame ou Monsieur), l’expression de mes sentiments distingués.

Michael P. Shea
Appendix E
Letter to Department Heads (English Version)

Michael P. Shea, M.A.
Department of Psychology
University of Windsor
401 Sunset Avenue
Windsor, Ontario
N9B 3P4

September 20, 1988

(Department Head's Name & Address)

Dear (Name):

As part of my doctoral dissertation in the Applied Social Psychology program at the University of Windsor, I am currently conducting a study of program evaluation utilization in Canada. My dissertation supervisor is Dr. Shelagh Towson.

I am attempting to locate as many Canadian program evaluators as possible to ensure that I have a representative sample of all Canadian program evaluators. As part of this search, I am writing to the department and program heads of all Canadian university departments with graduate training programs in criminology, education and educational administration, psychology, public administration, social work and sociology.

As the department or program head of a (Type of Department) department, I would greatly appreciate your assistance in this task. You can be of assistance by forwarding a copy of the enclosed survey package to faculty members or senior graduate students in your department or program who are involved in the conduct of Canadian program evaluation research.

If you would like further information on this study or require additional copies of the survey package please feel free to call or write. I would be happy to send you a summary of the results of this study should you wish. My telephone number is 519-253-4232, extension 2219. Thank you for your assistance with this matter.

Sincerely yours,

Michael P. Shea

401 Sunset Avenue, Windsor, Ontario, Canada N9B 3P4, 519/253-4232
Appendix F
Letter to Department Members (French Version)

Michael P. Shea
Department of Psychology
University of Windsor
Windsor, Ontario
N9B 3P4

Le 20 septembre, 1988

Monsieur/Madame,

Je fais actuellement un sondage sur l'utilisation de l'évaluation de programme au Canada. Je suis un candidat au doctorat en psychologie sociale appliquée à l'University of Windsor, travaillant sous la direction du Dr. Shelagh Towson. Depuis que j'ai mené ma première évaluation de programme il y a dix ans, je me suis intéressé aux facteurs qui influencent l'utilisation de l'évaluation de programme au Canada. Ce sondage, qui est le sujet de ma thèse de doctorat, est une enquête sur ces facteurs, et je crois que les données que j'espère obtenir seront d'intérêt pour tous les évaluateurs canadiens. Puis-je demander de participer à ce sondage en remplissant le formulaire ci-joint? Vous trouverez ci-inclus une enveloppe adressée et timbrée.

En tant qu'évaluateur de programme canadien, vous êtes tout spécialement qualifié pour fournir des renseignements sur ce sujet. Dans le but d'obtenir le plus grand échantillon possible d'évaluateurs canadiens, ce sondage a été envoyé à tous les départements d'universités canadiens qui offrent un enseignement supérieur en criminologie, enseignement, administration d'enseignement, administration publique, psychologie, service social, et en sociologie. Afin d'obtenir un échantillonnage complet, il est très important que tous les formulaires du sondage soient remplis et retournés.

Soyez assuré que vos réponses sont confidentiels. Votre identité, ainsi que celle de votre organisation resteront anonymes. Le formulaire porte un numéro d'identification uniquement pour que votre nom soit rayé de la liste d'envoi quand il est retourné. Seulement des données globales de l'échantillon seront rapportées.

Si vous desirez recevoir un sommaire des résultats, veuillez me l'indiquer sur la page où se trouve le formulaire d'acceptation.

Michael P. Shea

401 Sunset Avenue, Windsor, Ontario, Canada N9B 3P4, 519/253-4232
Appendix F
Letter to Department Members (English Version)

Michael P. Shea, M.A.
Department of Psychology
University of Windsor
401 Sunset Avenue
Windsor, Ontario
N9B 3P4

September 20, 1988

Dear Colleague:

I am a Ph.D. student in Applied Social Psychology at the University of Windsor being supervised by Dr. Shelagh Towson. Ever since I conducted my first program evaluation research study more than ten years ago, I have had questions about what factors influence program evaluation utilization in Canada. I believe that information on this issue is of import to other Canadian program evaluators.

As a member of a university department of (Subject’s University Department Type), you are uniquely qualified to supply information on this vital topic. I am attempting to contact all members of your department who have recently completed a Canadian program evaluation study through a request to your department head. To truly represent the population of Canadian program evaluators it is extremely important that each survey be completed and returned.

You may be assured of complete confidentiality. Your name and your department will remain anonymous. The survey has an identification number for mailing purposes only. This is so I may check your name off the mailing list when the survey is returned.

You may receive a summary of the results by indicating this preference on the consent form page of this survey which will removed from the survey upon receipt.

I or my dissertation supervisor, Dr. Shelagh Towson, would be most happy to answer any questions you might have. Please write or call. The telephone number is 519-253-4232, local 2250.

Thank you for your assistance.
Sincerely,

Michael P. Shea
Appendix G

METHODOLOGICAL NOTE #1

Response Rates

In an attempt to collect information on the amount and type of recent Canadian program evaluation utilization and on factors related to utilization success, a total of 1299 questionnaires were mailed to a sample of Canadian program evaluators on September 20, 1988, using two separate methods: the total design method advocated by Dillman (1978) and a university mailing.

Of the total number of 1299 questionnaires, 868 or 66.8% were personally addressed and mailed directly to all members of the Canadian Evaluation Society (CES) and to all members of the Program Evaluation Section of the Canadian Psychological Association (PES-CPA) through names and addresses obtained from the membership directories of these organizations. Of this group of 868 questionnaires, 764 or 88.0% were sent to potential subjects who were CES members only. An additional 84 or 9.7% were sent to potential subjects who were members of PES-CPA members and 20 (2.3%) were sent to potential subjects who belonged to both organizations.
Of these 868 personalized survey packages, 33 (3.8%) were returned unopened by the post office and subsequent follow-ups with CES and PES-CPA did not produce current addresses. The remaining 835 potential CES and PES-CPA research participants were followed up using survey methodology prescribed by Dillman (1978). Table G-1 shows the steps of the total design method of survey respondent follow-up with the returns obtained at each step. A total of 591 (70.8%) of these 835 questionnaires were returned by the intended respondents (as of July 1, 1989). The other 244 (29.2%) of the 835 CES and PES-CPA members whose questionnaires were not returned by Canada Post did not respond.

Of the 591 CES and PES-CPA survey respondents, a subtotal of 237 (40.1%) were willing to complete the questionnaire but indicated on a screening question that they had not conducted a Canadian program evaluation utilization that had been completed between January 1, 1985 and December 31, 1987. These CES and PES-CPA survey respondents did not complete the remainder of the survey and were eliminated from all further analyses with the exception of the non-respondent analyses. In addition, 15 or 2.5% of the 591 survey respondents returned incomplete questionnaires on which they indicated that they refused to participate in the study. This subgroup of survey respondents was also not included in subsequent data
Table G-1

**Total Design Method Response Rates by Response Category**

<table>
<thead>
<tr>
<th>Step</th>
<th>Date</th>
<th>Description</th>
<th>Response Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>September 20, 1988</td>
<td>Personalized Mailing</td>
<td>n = 278 out of 591 returns (47.0%)&lt;br&gt;278 out of 868 mailed (32.0%)</td>
</tr>
<tr>
<td>2</td>
<td>September 27, 1988</td>
<td>Reminder Card Mailing</td>
<td>n = 230 out of 591 returns (38.9%)&lt;br&gt;230 out of 868 mailed (26.5%)</td>
</tr>
<tr>
<td>3</td>
<td>October 25, 1988</td>
<td>Non-respondents' Second Package Mailing</td>
<td>n = 83 out of 591 returns (14.0%)&lt;br&gt;83 out of 868 mailed (9.6%)</td>
</tr>
</tbody>
</table>

**Notes.** Total Mailed (September 20, 1988) = 868. Total Returned (as of July 1, 1989) = 591 (68.1%).
analyses with the exception of the non-respondent analyses.

A subset of 339 (57.4%) of the 591 CES and PES-CPA respondents indicated that they had conducted a Canadian program evaluation study completed after January 1, 1985. Of this group, 292 (86.1%) were aware of how the evaluation findings had been utilized and completed the questionnaire. The remaining 47 survey respondents (13.9%) had conducted a Canadian evaluation during the selected time frame but were not aware of how their evaluation results had been utilized. This subgroup of survey respondents did not complete the remainder of the questionnaire and was therefore excluded from all but the non-respondent analyses.

The final yield for the total design method was 292 completed surveys, or 33.6% of the 868 surveys originally mailed in September of 1988. These 292 completed surveys represent 88.0% of the total sample of data used in the present study.

In an attempt to cast a wide net for Canadian program evaluators who might not belong to either the CES or the CPA Program Evaluation Section, an additional 431 questionnaires (33.2% of the total mailed) were sent to the department heads of 135 Canadian university departments. The frequencies of different kinds of responses received from these departments are listed in Table G-2. A minimum of three questionnaires were sent to each of the following
Table G-2

Response Frequencies by Type of University Department

<table>
<thead>
<tr>
<th>DEPARTMENT TYPE</th>
<th>n</th>
<th>R</th>
<th>NR</th>
<th>NO PE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Criminology</td>
<td>6</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Education &amp; Educ. Admin.</td>
<td>38</td>
<td>11</td>
<td>22</td>
<td>5</td>
</tr>
<tr>
<td>Psychology</td>
<td>34</td>
<td>6</td>
<td>16</td>
<td>12</td>
</tr>
<tr>
<td>Public Administration</td>
<td>16</td>
<td>4</td>
<td>10</td>
<td>2</td>
</tr>
<tr>
<td>Social Work</td>
<td>16</td>
<td>6</td>
<td>9</td>
<td>1</td>
</tr>
<tr>
<td>Sociology</td>
<td>29</td>
<td>10</td>
<td>13</td>
<td>6</td>
</tr>
</tbody>
</table>

Notes. (n) Number of departments contacted. (R) Number of departments returning questionnaires. (NR) Number of departments not returning questionnaires. (NO PE) Number of departments indicating no program evaluators in department.
types of university departments (number of departments in parentheses): Criminology (4), Education and Educational Administration (38), Psychology (32), Public Administration (16), Social Work (16) and Sociology (29), for a total of 405 questionnaires. The other 26 questionnaires were sent to Canadian university departments known to have relatively high numbers of program evaluators such as the Applied Social Psychology programs at the University of Saskatchewan and Wilfrid Laurier University and the program in measurement and evaluation at the Ontario Institute for Studies in Education. No additional follow-up procedures were conducted with this group of 431 questionnaires.

Only 89 (20.7%) of the 431 questionnaires sent to Canadian university departments were returned. It is not possible to determine how many of the non-returned questionnaires were actually given to university department members by department heads. However, department heads of 27 of the 135 university departments contacted sent a reply which indicated that no program evaluation research was being conducted by department members, accounting for an additional 81 (18.8%) of the 431 questionnaires. Of the remaining 350 questionnaires that may have been given to university department members by their department heads, 89 (25.4%) were returned.

Fifty (56.2%) of these 89 university-based respondents indicated that they had completed a Canadian program
evaluation study after January, 1985, and 40 of these 50 (80.0%) were aware of how the evaluation findings were utilized and completed the questionnaire. The other 39 (43.8%) indicated that they had not conducted an evaluation study during this time period despite the contrary belief of their department head.

The 40 completed surveys returned by university based evaluators represent 12.1% of the total data used for subsequent analyses. The total yield of usable responses from the university mailing was 40 out of 431 surveys mailed or 9.3%.

Response Rate Comparison by Method

Of the 1299 questionnaires mailed to potential study participants, a total of 680 (52.4%) were returned with 332 (25.6%) supplying data for the present study. Table G-3 presents a comparison of the return rates obtained from the use of the two different methods as well as the combined return rate totals.

In the combined group of CES, PES-CPA and university respondents, 332 of the 680 (48.8%) had completed a Canadian program evaluation study after January 1, 1985, were aware of how the results had been utilized and completed the questionnaire. Their data formed the basis
Table G-3

Response Categories by Survey Method

<table>
<thead>
<tr>
<th>METHOD</th>
<th>TOTAL</th>
<th>RC</th>
<th>RIU</th>
<th>RIN</th>
<th>RR</th>
<th>RU</th>
<th>NR</th>
</tr>
</thead>
<tbody>
<tr>
<td>n</td>
<td>591</td>
<td>292</td>
<td>47</td>
<td>237</td>
<td>15</td>
<td>33</td>
<td>243</td>
</tr>
<tr>
<td>%</td>
<td>68.1</td>
<td>33.5</td>
<td>5.4</td>
<td>27.3</td>
<td>1.7</td>
<td>3.8</td>
<td>28.1</td>
</tr>
</tbody>
</table>

Univ. Mailing  
(n = 431)  
| n            | 89    | 40  | 10  | 39  | 0   | 0   | 342 |
| %            | 20.7  | 9.3 | 2.3 | 9.1 | 0.0 | 0.0 | 79.4|

Combined     
(N = 1299)  
| n            | 680   | 332 | 57  | 276 | 15  | 33  | 585 |
| %            | 52.4  | 255.5| 4.4 | 21.2| 1.2 | 2.5 | 45.0|

**Notes.** (RC) Returned completed questionnaire, had completed a recent Canadian program evaluation and was aware of how it was utilized. (RIU) Returned incomplete questionnaire, had completed a recent Canadian program evaluation but unaware of how it was utilized. (RIN) Returned incomplete questionnaire, had not completed a recent Canadian program evaluation. (RR) Returned incomplete questionnaire, refused to participate. (RU) Questionnaire returned undelivered by Canada Post. (NR) No response.
for this study. The vast majority of other respondents had not conducted a Canadian evaluation project during this time period despite their membership in the two evaluation organizations or the belief of their university department head that they had.

In comparing the yields of the two methods, the survey methodology of Dillman (1978) clearly produced a much higher response rate. However, it not known what percentage of the 431 questionnaires sent to university department heads were actually given to program evaluators. In addition, it should be noted that 30 of the CES and PES-CPA respondents were also university faculty members. This resulted in a reduction of the total number of potential university respondents.
Appendix H

METHODOLOGICAL NOTE #2

Non-Respondent Analyses

Only 590 of the 868 (68.0%) questionnaires sent to CES and PES-CPA members were returned. In an attempt to determine whether this sample of CES and PES-CPA members truly represented the total membership populations, the 278 CES and PES-CPA non-respondents and 590 respondents were compared on four characteristics derived from the CES and PES-CPA membership directories: organizational membership, language preference, gender, and work setting. These data are presented in Table H-1.

Based on the comparison of these four characteristics, CES and PES-CPA non-respondents do not differ significantly from the respondents. Therefore, it can be argued that the respondents, who make up 68.0% of the total group of CES and PES-CPA members, are a representative sample of the total membership of these two organizations.

It was not possible to compare university respondents with the non-respondents from this group due to a lack of data on non-respondents.
### Table H-1

**Comparison of CES and PES-CPA Respondents and Nonrespondents**

<table>
<thead>
<tr>
<th></th>
<th>n</th>
<th>Respondents</th>
<th>Non-Respondents</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>368</td>
<td>590 (68.0%)</td>
<td>278 (32.0%)</td>
</tr>
<tr>
<td><strong>Organization:</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CES only</td>
<td>764</td>
<td>519 (88.0%)</td>
<td>245 (88.1%)</td>
</tr>
<tr>
<td>PES-CPA only</td>
<td>84</td>
<td>53 (9.0%)</td>
<td>31 (11.2%)</td>
</tr>
<tr>
<td>CES &amp; PES-CPA</td>
<td>20</td>
<td>18 (3.1%)</td>
<td>2 (0.7%)</td>
</tr>
<tr>
<td><strong>Language Preference:</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>French</td>
<td>108</td>
<td>70 (11.9%)</td>
<td>38 (13.7%)</td>
</tr>
<tr>
<td>English</td>
<td>760</td>
<td>520 (88.1%)</td>
<td>240 (86.3%)</td>
</tr>
<tr>
<td><strong>Gender:</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>454</td>
<td>318 (53.9%)</td>
<td>136 (48.9%)</td>
</tr>
<tr>
<td>Female</td>
<td>280</td>
<td>184 (31.2%)</td>
<td>96 (34.5%)</td>
</tr>
<tr>
<td>Unknown</td>
<td>132</td>
<td>88 (15.1%)</td>
<td>46 (16.1%)</td>
</tr>
<tr>
<td><strong>Work Setting:</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Federal Govt.</td>
<td>188</td>
<td>125 (21.1%)</td>
<td>63 (22.7%)</td>
</tr>
<tr>
<td>Provincial Govt.</td>
<td>203</td>
<td>144 (24.7%)</td>
<td>59 (20.6%)</td>
</tr>
<tr>
<td>Not For Profits</td>
<td>43</td>
<td>29 (5.0%)</td>
<td>14 (4.9%)</td>
</tr>
<tr>
<td>University</td>
<td>89</td>
<td>58 (10.0%)</td>
<td>31 (10.8%)</td>
</tr>
<tr>
<td>Consulting Co.</td>
<td>95</td>
<td>58 (10.0%)</td>
<td>37 (12.9%)</td>
</tr>
<tr>
<td>Unknown</td>
<td>234</td>
<td>157 (26.9%)</td>
<td>77 (26.9%)</td>
</tr>
</tbody>
</table>
VITA AUCTORIS

Michael Patrick Shea was born on July 5, 1953 in Toronto, Ontario to Harold Cecil Shea and Margaret Mary Shea (nee Wallace). He attended Our Lady of Grace and St. Joseph's Elementary Schools in Aurora, Ontario. In 1966, he moved to Glendale, California with his mother and new stepfather, Steven Sagodi. In June 1971, he graduated from Glendale High School as a member of the National Honor Society. After returning to Canada in the summer of 1971, Michael attended Vancouver Community College's Langara campus prior to enrolling at Simon Fraser University. He graduated with a Bachelor of Arts (Honours) degree in Psychology in June, 1976. He worked as a youth worker for a year and then returned to S.F.U. During his graduate work at S.F.U., Michael directed a major program evaluation study for the S.F.U. Criminology Research Centre. In June of 1982, he was awarded a Master's degree in the area of Applied/Clinical Psychology program. After leaving S.F.U., Michael was employed as the Director of the Coquitlam Family Centre until he enrolled in the Ph.D. program in Applied Social Psychology at the University of Windsor in September, 1985. Michael is currently employed as a Research Associate by the Centre for Research in Human Development at Laurentian University.

Michael is married to the former Alexandra Lesia Kroitor. They have a daughter named Stephanie Margaret.