UNIVERSITÉ DU QUÉBEC À MONTRÉAL

LA DISTINCTION ENTRE LE PROCESSUS SPÉCIFIQUE DE L'ÉVALUATION DE PROGRAMME ET L'APPLICATION DE SON INSTRUMENTATION : UNE ÉTUDE DE CAS DES MODÈLES EN ÉVALUATION DE PROGRAMME

MÉMOIRE
PRÉSENTÉ
COMME EXIGENCE PARTIELLE
DE LA MAÎTRISE EN ÉDUCATION

PAR
LINA ZIELINSKI

JUILLET 2007

UNIVERSITÉ DU QUÉBEC À MONTRÉAL

DISTINGUISHING BETWEEN THE SPECIFIC PROCESS OF PROGRAM
EVALUATION AND THE APPLICATION OF ITS INSTRUMENTATION: A
CASE STUDY OF PROGRAM EVALUATION MODELS

THESIS
PRESENTED
IN PARTIAL FULLFILMENT OF THE REQUIREMENTS
FOR THE MASTER OF EDUCATION

BY
LINA ZIELINSKI

JULY 2007

UNIVERSITÉ DU QUÉBEC À MONTRÉAL
Service des bibliothèques

Avertissement

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my thesis supervisor Ms. Marthe Hurteau for her help and guidance. Her devotion to her students is unequalled, her passion for research is infectious and her knowledge of the field of evaluation is inspiring. Over and above her own research, presentations and teaching task she has always made herself available to answer my needs and ensure the timely completion of each phase of my thesis. I feel honored and privileged to have worked with her. I would also like to thank the members of my jury M. Sylvain Houle and Ms. Marjolaine Saint-Pierre for having taken the time to read my work and for having shared their insight. They provided valuable information which I've integrated in my thesis.

Words cannot express the gratitude I feel towards my family. Without their love, support, compassion and understanding, I would not have been able to complete such an endeavour. Although I often gave priority to my thesis, they created an environment in which I could work, free of guilt. Their selflessness and patience humble me.

I would also like to thank Mr. Brad Cousins, at the University of Ottawa, who had supervised my work at the onset of my graduate studies. As well, I appreciate the insight provided by Ms. Marielle Simon and Mr. Tim Aubry who were members of the jury when I presented my original proposal at the University of Ottawa. I truly appreciate the time and effort they so generously offered during the first phase of my graduate work.

I would like to acknowledge Suzanne Boyer who helped me develop time management skills. Her insight allowed me to balance my personal life, my professional career and my studies; for this I am truly grateful.

Finally, I dedicate this thesis to the memory of my father and I thank God for the perseverance required in completing this endeavour.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

RÉSUMÉ

Maintenant que le Programme de formation de l'école québécoise est implanté depuis plusieurs années au niveau primaire, il fait l'objet d'évaluation. Par contre, le domaine de l'évaluation de programme se fait souvent remettre en question par rapport à sa capacité de produire des recommandations pertinentes. Il existe donc un besoin de recherche fondamentale dans le domaine de l'évaluation de programme. Pour améliorer l'évaluation de programme dans le domaine de l'éducation, il est essentiel d'effectuer de la recherche fondamentale dans le domaine de l'évaluation de programme.

Fournier (1995) et Hurteau et Houle (2006) ont effectué des recherches qui visaient la pratique en évaluation de programme. Leurs réflexions portaient sur la distinction entre ce qui constitue la pratique et ce qui a été élaboré pour la soutenir. Par contre, il existe une confusion dans la littérature à ce niveau car cette distinction n'a jamais fait l'objet d'une étude empirique. Notre recherche s'inscrit dans cette perspective afin d'établir un cadre de référence pertinent et d'examiner si le processus propre à l'évaluation de programme diffère de l'application de son instrumentation.

Nous avons analysé des réponses à un scénario simulé. Ces réponses représentent trois orientations théoriques : méthodes, jugement de valeur et utilisation. Même si ces orientations ne sont pas nécessairement des catégories mutuellement exclusives, chacune s'attarde à une perspective distincte. L'orientation théorique associée aux méthodes utilise une approche expérimentale ou quasi-expérimentale où la rigueur est fondamentale. L'orientation du jugement de valeur souligne l'importance de porter un jugement sur la valeur du programme évalué afin d'offrir l'information nécessaire à une prise de décision. L'orientation théorique de l'utilisation s'attarde à l'utilisation des résultats de l'évaluation afin d'obtenir un changement au niveau organisationnel.

Des grilles d'analyse sont utilisées pour identifier les éléments du processus de l'évaluation de programme et une analyse croisée des données quantitatives démontre le taux de présence de chaque élément dans chaque orientation théorique. Ces données sont ensuite appliquées à une analyse de contenue et font l'objet d'une analyse qualitative. Les composantes de l'instrumentation sont aussi appliquées à chaque orientation théorique. Les standards de pratique provenant du Joint Committee servent de cadre de référence pour identifier les variances et les choix méthodologiques sont étudiés à partir de critères prédéterminés.

Les résultats démontrent qu'à un niveau global les trios orientations théoriques ont présenté les éléments du processus de l'évaluation de programme. Par contre, des différences ont été identifiées dans l'orientation de l'utilisation. C'est à souhaiter que la présente recherche éclaire notre sujet en stimulant la discussion et de futures études.

ABSTRACT

Now that the Quebec Education Program has been implemented, evaluators are being called upon to assess its contributions. However, the field of program evaluation's ability to provide relevant recommendations has been said to be questionable. As such, there exists a desperate need for fundamental research in the field of program evaluation. It stands to reason that in order to improve program evaluation in the field of education, we must conduct fundamental research in the field of program evaluation.

Researchers such as Fournier (1995) and Hurteau & Houle (2006) have provided some insight into program evaluation practices. These researchers distinguish between that which constitutes the practice and that which was elaborated to support it. However, there seems to exist confusion in the literature regarding this distinction since it has not been empirically studied as of yet. It is within this perspective that our study's main goal is to examine whether the program evaluation process differs from the application of its instrumentation so as to establish a relevant frame of reference.

In order to conduct such an investigation, we first analyze responses to a simulated scenario presented from three different theoretical perspectives, valuing, methods and use. Although each orientation is not necessarily a mutually exclusive category, their main focus differs from one orientation to the next. The main focus of the valuing theoretical orientation is on providing a judgment of the program's merit or worth in order to inform the decision making process. The main concern of the methods theoretical orientation main is with the use of experimental or quasi-experimental methodology in order to augment the evaluation's rigor. According to the use theoretical orientation, the ultimate goal of program evaluation is for the results to be used in order to achieve organizational change so as to best serve the client.

Analysis of the simulated scenario responses are presented in an analysis grid which shows identified elements of the program evaluation process and a quantitative cross-analysis of the data is performed to show how this data pertains to the rate of occurrence of each element, in each theoretical orientation. Then, this data is applied to a content analysis and becomes the object for qualitative investigation. Components of the instrumentation are applied to each theoretical orientation. The Joint Committee's Program evaluation Standards serve as a frame of reference in identifying discrepancies within each theoretical orientation and criteria are formulated to guide the investigation regarding the methodological choices.

The results obtained in our research show that, globally, the three theoretical orientations did indeed present the elements of the program evaluation process, with discrepancies identified in the use orientation. Nonetheless, the goal is not to trivialize the value of any given theoretical orientation; it is to shed light on the topic in order to stimulate discussion and future research.

INTRODUCTION

Since the 1960's the field of program evaluation has been growing and expanding. With curriculum reforms at the forefront of on going debates in the field of education, there exists a growing need for evaluations to be conducted in this area so as to provide relevant findings in order to address the changing needs in society. However, program evaluation's ability to generate relevant findings which enlighten the decision-making process is regularly called into question. As such, conducting fundamental research in the field of program evaluation with the goal of improving current practices, will inevitably improve program evaluation in the field of education. In an attempt to guide evaluators in their practices, researchers from various theoretical orientations have developed evaluation models which created a certain confusion since practitioners commonly associate the application of the models with the practice of program evaluation. Through her research on the application of evaluation models, Christie (2003) reached the following conclusion: "Over the past thirty years eminent evaluation theorists have appealed for increased empirical knowledge of evaluation, based on the notion that such knowledge is necessary to explain the nature of evaluation practice ... Although an appeal for the empirical study of evaluation practice has been made repeatedly, it has met with little response" (p.7).

Recent reflections emanating from Hurteau and Houle's (2006) research, have allowed these authors to establish a distinction within the practice (as suggested by Christie, 2003); the specific act of evaluating and the instrumentation elaborated to support this act (evaluations models being a prime example). However, this distinction has never been empirically studied and so this research is an effort to examine the distinction between the practice and the instrumentation which supports it. This distinction paves the way for research to move beyond by questioning whether models fulfill their supportive function. It is within this perspective that the

present research attempts to examine whether the program evaluation process differs from the application of its instrumentation.

The intent is neither to diminish the value of evaluation models nor to trivialize them but rather to increase the body of knowledge surrounding evaluation practice in order to improve the practice which is, as mentioned, often called into question.

The first chapter, the problem statement, will provide the fundamental information needed to contextualize the present study. Emphasis will be placed on the problem arising from an existing gap in evaluation literature. This chapter will build a foundation on which the main research will rest.

A review of literature will constitute the second chapter. This chapter will present a working definition of program evaluation and the instrumentation. Particular emphasis will be placed on evaluation models developed from different theoretical perspectives since they will be the main component of the instrumentation that will be investigated.

The methodology will be presented in the third chapter. The research's intended design, the source of data, the instrumentation, the plan of analysis and its rigor will constitute this third chapter.

The fourth chapter will present and analyze the results emanating from the collected data. It will do so by providing summary tables of the quantitative findings and direct quotes to support the qualitative findings.

The final chapter of the present study is the discussion which will highlight key findings in an attempt to shed light on the studied dilemma.

CHAPTER I

STATEMENT OF PROBLEM

This first chapter serves to contextualize and state the problem of the present research. Our main focus is on the program evaluation process and the application of evaluation models in this process. The reflection that arose from these considerations enabled us to formulate a problem statement as well as the main research hypothesis.

1. DEFINING PROGRAM EVALUATION

Lincoln (1985) elaborated her own definition of program evaluation which introduced terms such as "controlled investigation" and "value of the unit as a whole". Scriven (1991) subscribes to this perspective and emphasizes that program evaluation is based on determining the value or the merit of a unit. McLaughlin and Jordan (2004) add that evaluation is: "An intentional transformation of specific resources (inputs) into certain activities (processes) to produce desired outcomes (results) within a specific context" (p. 7).

Thus, program evaluation involves a systematic gathering of data whose main objective is to study feedback systems in order to render a value judgement, so as to improve or to influence decisions concerning its future outcome (Patton, 2002; Patton, in Alkin, 1990; Scriven, 1991; Stake, 2004). It is part of a program management perspective. In order to better understand the nature of its goals, due consideration must be given to its target audience as well as to the very nature of the judgements it generates.

Stake (2004) summarizes the positions held by establishing that evaluations offer an assessment of the value and performance of a program so as to improve it. Mathison (2004) defines program evaluation as being a rigorous process which involves rendering a judgement regarding a program's value and is part of sound management practices. Although these definitions do allow for a general portrait of program evaluation, evidently, there exist numerous definitions of program evaluation and no single one has taken hold among all evaluators (Mark, Greene & Shaw, 2006).

2. THE EVOLUTION OF THE FIELD OF PROGRAM EVALUATION

Although program evaluation is commonly thought to be a recent phenomenon, it originated over 150 years ago. This misconception may be explained in large part by the fact that it wasn't until the end of the 1960's that program evaluation received considerable attention from the scientific community. It is during this period that researchers in the field of social sciences were called upon to evaluate the efficiency of numerous programs that were highly acclaimed following the Second World War (Madaus and Stufflebeam, 2004). Although originally program evaluation applied social science research techniques to render a value judgment regarding a program's performance (Suchman, 1967), it became, with time, a separate discipline that has distinguished itself from research and evaluative research by its unique ability to provide an effective management tool aimed at enlightening the decision-making process (Hurteau & Houle, 2005b).

The 1960's brought a "boom" in program reforms and with it an increasing need for a new perspective of evaluation. Nonetheless, the trend in program evaluation grew and gained strength. Evaluation specialists were sought out and called upon more frequently and diversified methods in evaluation became more apparent. As a result, a

growing need to define a process unique to the field of evaluation arose. In response, evaluation theorists focused their efforts on elaborating evaluations models and methods that would help guide the practice. Thus, the 1970's and 1980's gave rise to professionalizing the field of evaluation and research attempted to structure the evaluation process by proposing various methods and models (Madaus and Stufflebeam, 2004). As a result, it was during this period that prominent members of the evaluation community united to form The Joint Committee on Program Evaluation. These members elaborated a set of standards intended to guide evaluators in the practice of rendering the system accountable and evaluating new programs' level of efficiency.

As was experienced in the 1980's, the field continued to expand in the 1990's. However, a growing concern within the evaluation community regarding the use of evaluation findings began to surface; recommendations made in evaluation reports were not being implemented as intended and/or evaluation findings were being misused. As a result, this period was characterized by an emphasis on models oriented toward the client. Stufflebeam's (1983, 2000) Context-Input-Process-Product (CIPP) Model, Patton's (1978, 1986, 1997) Utilization-Focused Evaluation (UFE) and Cousins' (1992, 1998) Participatory Evaluation (PE) attempted to rectify the dilemma by integrating stakeholder participation in the evaluation process.

Although research in the 1990's focused on problems encountered in the practice, research in the new millennium is shifting toward causal elements and solutions. Stake's (2004) contribution to the field of evaluation is undeniable since he emphasizes the significant value of criteria and standards as being fundamental components in all evaluations given that, without them, this activity remains at a descriptive level, rendering the capacity to make value judgment very difficult.

Finally, since much time, effort and resources have been spent on developing and revising standards of practice and evaluation models, the new millennium has brought with it a willingness to consolidate the acquired knowledge through accredited training programs offered by professional associations. Thus, research on evaluation is beginning to focus on the practice's fundamental elements and components.

3. THE PROGRAM EVALUATION PROCESS AND THE INSTRUMENTATION

The evolution which occurred within the field of practice, the confusion which persists between the different constituents in evaluative practices as well as recent studies have lead Hurteau & Houle (2006) to establish a distinction between that which constitutes the specific act of evaluating and the different elements which contribute to its support. Hurteau & Houle (2006) refer respectively to the modeling of the program evaluation process, and the instrumentation. The figure presented on the following page illustrates this concept.

In essence, the modeling of the program evaluation process is central since it establishes the components that are associated to the specific act of evaluating which distinguishes it from all other actions. Hurteau & Houle (2006) present the instrumentation in four components; the Joint Committee's standards of practice, program evaluation models/theoretical orientations, methodological choices and ethical guidelines offered by different evaluation associations. The instrumentation focuses on *how* the evaluation will be conducted.

Figure 2.1: The modeling of the program evaluation process and the instrumentation
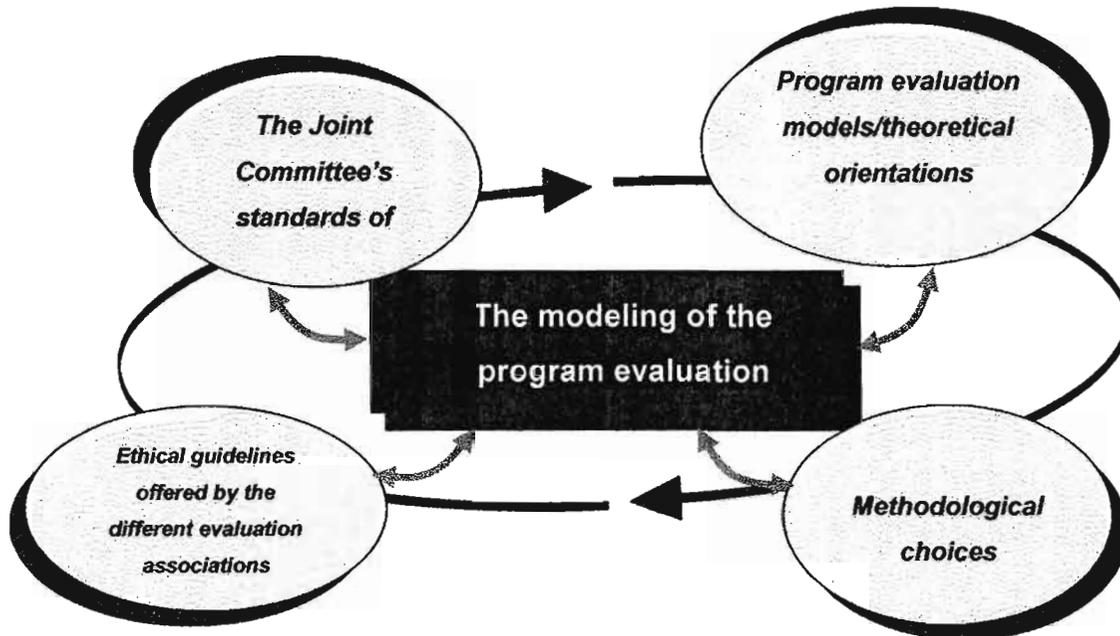
Hurteau & Houle, 2006

The following sections will present an overview of each aspect of Figure 2.1 and will be further developed in the following chapter.

3.1 The modeling of the program evaluation process

Originally, Scriven (1980) proposed the concept of the Logic of Evaluation which involves the following four stages: (1) establishing criteria; (2) elaborating standards; (3) measuring performance in relation to the established standards and; (4) synthesizing and integrating the data to render a judgment. He later referred to the phrase Logic of Evaluation as being "the specific principles of reasoning that underlie the inference processes in all and only the fields of evaluation" (Scriven, 1995). Thus, he associates the practice of evaluation to the Logic of Evaluation. Some authors, such as Hurteau (1991), Fournier (1995) and Stake (2004) pursued Scriven's original

thinking and the culmination of their reflections lead to an elaboration of the modeling of the program evaluation process (Hurteau & Houle, 2005a; Hurteau, Houle & Duval, 2005). According to these authors, this modeling is involved in all evaluations and synthesizes the different perspectives. The modeling of the program evaluation process has been validated (Hurteau, Lachapelle & Houle, 2006) and it includes the following 6 elements: (1) description of the program to be evaluated; (2) establishing the evaluation's dilemma at the onset of the evaluation; (3) rendering the evaluation operational; (4) strategic choices; (5) making evaluative claims and; (6) synthesizing/evaluative conclusions.

This reflection will be presented in the following chapter. For the purposes at hand, suffice it to say that program evaluation practice follows a logical set of specific activities resulting in a value judgment that is based upon identified objectives and standards established in the early phases of the process. Thus, evaluation practice is the specific act of evaluating and, as such, has a specific set of operations that are unique to the practice of evaluating which are presented in the modeling of the program evaluation process. In this sense, the modeling of the program evaluation process is the specific act of evaluating and is a generic representation of the practice (Hurteau & Houle, 2006).

3.2 The instrumentation

The instrumentation is considered to be that which supports the modeling of the program evaluation process. Hurteau & Houle (2006) have identified standards of practice, ethical guidelines, methodological choices and evaluation models as the different aspects of the instrumentation.

In an attempt to guide the evaluator, the Joint Committee (1981) developed Standards for Evaluation of Educational Programs, Projects and Materials which were revised in 1994. The Program Evaluation Standards are a set of 30 standards divided into four categories: utility, feasibility, propriety and accuracy. Also, every major national evaluation society developed its own ethical guidelines which are all very similar in nature. For example, the Canadian Evaluation Society (CES) produced Guidelines for Ethical Conduct with three main guidelines (competence, integrity and accountability) and 11 defining statements.

Finally, various evaluation models from different theoretical orientations were elaborated. The past decades proposed so many models elaborated from so many different perspectives that it became necessary to organize them. Alkin and Christie (2003) used a category system inquiry approach and proposed an Evaluation Theory Tree which served as a metaphor to illustrate their findings. These authors identified three major theoretical perspectives or orientations: the methods orientation, the valuing orientation and the use orientation. These theoretical orientations will be presented in the following chapter. For the purposes at hand, suffice it to say that Alkin and Christie (2003) chose to categorize almost 30 theorists and there are a countless number who were not included in the study. From this, it is safe to conclude that much time, effort and resources have been spent on elaborating evaluation models.

4. THE SITUATION'S CURRENT STATE

To date, studies that have been conducted and provide confirmation regarding the weak relationship between theoretical concepts and their application (Christie, 2003a; Datta, 2003; Chandler, 2001). Chandler (2001) found that: « ...evaluation theory did not overly influence the way most approached their practice » (p.3) and Christie's

(2003a) study supported this notions in that her findings confirm that practitioners rarely utilize program evaluation models and, in the isolated cases where they do use models, they sometimes have a tendency to denature them. Christie's (2003) investigation involved evaluators by asking them whether they use evaluation models to guide their practice. She found that less than 50% of the evaluators surveyed reported that they did not use a specific program evaluation model. These results prompted Christie draw the following conclusion:

> "It is more accurate to say that evaluators use some of the theoretical notions put forth by particular theories rather than fully implement a particular theory. Furthermore, the practices of those who did claim to be using a specific theory did not correspond empirically with the practices of the identified theorist. Thus, I conclude that the gap between the "common" evaluator and the notions of evaluation practice put forth by academic theorists has yet to be bridged" (pp.33-34)

Christie's findings were corroborated by Hurteau, Houle & Duval (2005) and Hurteau & Houle (2005a). Their study, which involved the analysis of 69 published evaluation articles, found that only 6% of the authors referred to evaluation models. This finding supports Christie's (2003) findings that evaluation models are rarely used by practitioners. Christie provides the following insight:

> ... it seems important to investigate the implicit or folk theories of evaluation that exist in the field, and how everyday practitioners form notions about evaluation and then use them to guide their work, then what does? Research on evaluation can help answer this question.
>
> (Christie, 2003b, p.92)

However, literature does provide some guidance regarding Christie's concerns. As mentioned (in section 3.1), Scriven (1980) proposed the concept of the Logic of Evaluation which involves the following four stages: (1) establishing criteria; (2) elaborating standards; (3) measuring performance in relation to the established standards and; (4) synthesizing and integrating the data to render a judgment. This

outlines the essential operations which should guide evaluators to formulate an adequate judgment which is the prime motive of all evaluations. Furthermore, Fournier (1995) share this perspective by distinguishing two types of logic that guide evaluation practice; the General Logic and the Working Logic that are distinct and, at the same time, in interrelation. Fournier's (1995), General Logic follows Scriven's four steps whereas the Working Logic focuses on the methods used to establish each of the four steps. As such, the General Logic transcends the structures proposed by different approaches (or models). This author supports her position by demonstrating the way in which these two logics apply differently contingent upon the evaluation models. Not only do her explanations establish a distinction between the two logics but she also presents them as being a generic process which takes on different forms of application depending on the models. As such, they are interrelated. The following figure illustrates this concept:

Figure 2.2. The relationship between Working Logic and General Logic
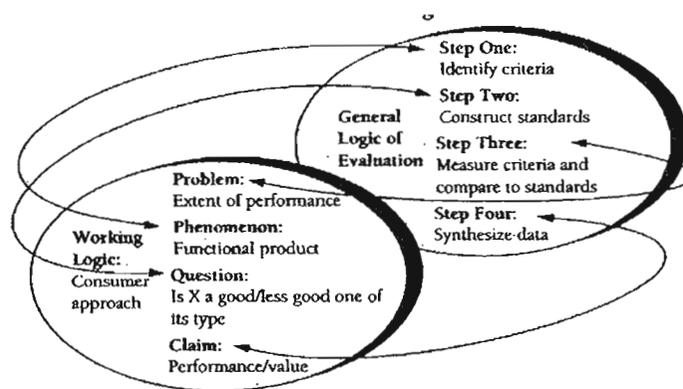
(Fournier, 1995)

In line with this perspective, is the distinction proposed by Hurteau & Houle (2006) regarding the specific process of program evaluation and the instrumentation. The

modeling of the evaluation process (validated by Hurteau, Lachapelle & Houle, 2006) presents the fundamental element inherent to the program evaluation process. The instrumentation refers to components that were elaborated to support the program evaluation process (including the evaluation models). Although the modeling of the program evaluation process distinguishes itself from the instrumentation, the two are presented as being interrelated (Figure 2.1).

The authors presented offer a generic process that translates the specific act of evaluating which should allow practitioners to render a credible judgment on the program being evaluated. They are all founded on the Logic of Evaluation (Scriven, 1980) and both Fournier (1995) and Hurteau & Houle (2006) distinguish it from the instrumentation yet show how they are interelated.

These assertions lead us back to Christie (2003b) in the sense that literature does offer interesting alternatives to evaluation models that can be used as a frame of reference to guide evaluators in their practice. However, attempts at this distinction remain at a reflective or theoretical level and perhaps it is for this reason that there exists, in the literature, an apparent confusion which calls for clarification between that which fundamentally constitutes evaluation – the specific act of evaluating – and its instrumentation – that which supports this act.

5. THE PROBLEM STATEMENT

Over the past thirty years, much resource has been deployed to develop methods that would serve as a beacon to practices in program evaluation. The development of standards, their revision and the elaboration of a multitude of models all follow in this direction. Other research made attempts to identify parameters that served as a generic outline of the practice.

Christie (2003) confirms "over the past thirty years eminent evaluation theorists have appealed for increased empirical knowledge of evaluation, based on the notion that such knowledge is necessary to explain the nature of evaluation practice ... Although an appeal for the empirical study of evaluation practice has been made repeatedly, it has met with little response."

Among these efforts, we can document numerous writings and developments that focused on evaluation models. Perhaps all this energy invested has been misplaced since recent studies (Christie 2003, Hurteau & Houle 2005a) have provided some insight on the fact that models are not being used in practice. This emphasizes a need for a frame of reference that is more relevant. Scriven (1980), Fournier (1995), Hurteau (1991) and Hurteau & Houle (2006) made an effort to distinguish between the act of evaluating and that which supports it and to offer a generic process that can be applied to all forms of evaluation. These reflections have brought about a distinction between the evaluation process (which is generic since it can be applied to all approaches in evaluation) and evaluation models which are more specific to the instrumentation (Hurteau & Houle, 2006; Fournier, 1995). This distinction could also respond, at least in part, to Christie (2003b) who asked what was being used in practice if not evaluation models.

As such, it becomes important to elaborate on this distinction so as to verify whether it is a plausible one. This would allow research to move beyond the evaluation models by refocusing the attention on the nature of the practice (Christie, 2003) and by revisiting fundamental concepts such as Scriven's (1980) Logic of Evaluation, the two logics proposed by Fournier (1995) and the process proposed by Hurteau & Houle (2006). As such, the purpose of this study is to examine this distinction and to investigate whether this distinction contributes to the practice in a relevant way. In this sense, we will gain a greater understanding of the nature of evaluation practice in the hopes of increasing best practices.

6. CONTRIBUTIONS

Program evaluation's ultimate goal is to provide valid and useful information rigorously gathered so as to offer clarity to the decision-making process. Considering the consequences involved in decision-making, it is of value to question the process that leads an evaluator to make his or her claims. As such, the present study is both scientifically and socially relevant.

On a scientific level, distinguishing between program evaluation models and the modeling of the evaluation process is a new point of view. Although research findings confirm that evaluation models are not being used, that which is being used is still being questioned. Examining influential elements would provide empirical data and could offer valuable information toward the elaboration of a synthesized core body of knowledge thus aiding in the development of program evaluation as a discipline. Dubois and Marceau (2005) contend that the primary difficulty encountered in the field which inhibits the development of program evaluation as a discipline in and of itself is that it is not founded on a core body of knowledge which harmonizes theory, methodology and empirical data. Since program evaluation is constantly questioned on its ability to generate relevant recommendations, it is evident that this problem persists.

On a social level, program evaluation is central to the educational systems in place in North-America. The field of education regularly calls upon evaluators to assess programs at different stages of their life cycle. Evaluation ensures regular feedback on the different curriculum aspects in existence and provides clarity to the decision-making process. Now that the Quebec Educational Plan has been implemented, evaluators are being called upon to assess its contribution. The consequences associated with these decisions require careful consideration attainable only with the use of rigorous techniques. These techniques involve a logical structure common to

all evaluations and allow for valid conclusions that respect a certain social justice. However, the field of program evaluation's ability to provide relevant recommendations has been said to be questionable.

Errors in judgments have been committed along the way. Prime examples of such errors include the evaluation of the *Follow Through* program in the United States and the *Opération Renouveau* program here in Montreal. Since both evaluations were based on identical criteria and indicators, the findings presented in the evaluations showed that the desired results of the program were not attained even though the milieus offering the program reported an improvement in the rate of delinquency. In order to avoid such errors in judgment, fundamental research in the field of program evaluation is of essence. As such, there exists a desperate need for fundamental research in the field of program evaluation. It stands to reason that in order to improve program evaluation in the field of education, we must conduct fundamental research in the field of program evaluation. The recognition of this responsibility inherent in program evaluation bestows upon it an undeniable social relevance.

Finally, although our study is limited to three theoretical orientations, its main goal is to examine whether the program evaluation process differs from the application of the instrumentation so as to gain a greater understanding of the nature of evaluation practice. It is hoped that this study will increase the body of knowledge surrounding evaluation practice by stimulating discussion in order to motivate further inquiry on the topic.

CHAPTER II

REVIEW OF LITERATURE

The problem raised in the previous chapter refers to the confusion that seems to exist in the literature regarding the distinction between the program evaluation process and the application of its instrumentation. The purpose of our study seeks to address this problem by considering the modeling of the program evaluation process which illustrates the specific act of evaluating and evaluation models that have been developed to support this act. Within this perspective, the present chapter will focus on presenting the modeling of the program evaluation process and the theoretical roots that stimulated the different orientations upon which current evaluation models are founded. Our study also investigates whether the distinction between the modeling of the evaluation process and program evaluation models is relevant to the practice. As such, this chapter will present the components of the instrumentation elaborated to support the practice. This will allow for precisions to be made regarding the research hypothesis and provide a solid methodological foundation upon which the present study rests.

1. THE MODELING OF THE PROGRAM EVALUATION PROCESS

The present section will define the practice of program evaluation by presenting its fundamental aspects and subsequent developments so as to provide a current definition.

1.1 Basic fundaments

Scriven (1995) mentions that evaluation is a form of inquiry and, by this notion, it needs a legitimate basis on which to found its claims if they are to be considered valid. Similarly, Taylor (1961), Toulmin (1964, 1984), McCarthy (1973, 1979), and Redding (1989) have all emphasized that any form of scientific inquiry possesses a general reasoning pattern which is interdisciplinary and is used to inform and guide practice. This "general pattern" is a fundamental reasoning supra-structure that legitimizes claims. Scriven's (1980) Logic of Evaluation provides this supra-structure to program evaluation which distinguishes it from any other form of inquiry.

As mentioned in the previous chapter, the concept of Scriven's (1980) Logic of Evaluation is comprised of the following four basic operations:

1. Selecting criteria of merit: which elements or components influence the performance of the object being studied (evaluand)?

2. Setting standards of performance on those criteria: which is the anticipated level of performance?

3. Gathering data pertaining to the evaluand's performance on the criteria relative to the standards (analysis): to which extent did the object perform in response to the standards?

4. Integrating the results into a final value judgment (synthesis): what is the value of the studied object?

These operations contain evaluation's fundamental parameters and processes; the criteria, the standards, the measures or observable variables and the judgment. These parameters transcend the unique nature of any individual evaluation and, as such, provide a foundation to all evaluations and validate their claims.

However, the four basic operations proposed by Scriven (1980) must be integrated so as to be of value and useful to the evaluation process. Some authors, (i.e. Hurteau, 1991; Fournier, 1995; Stake, 2004), have pursued Scriven's (1980) original reflection so as to, on the one hand, identify the parameters necessary in integrating the operations into an functional framework for the practice and, on the other hand, to consider the conceptual and methodological developments that have occurred since Scriven's proposed definition.

Hurteau (1991) emphasized the distinction between strategic choices and methodological choices. Strategic choices involve identifying and defining the object of evaluation, setting the criteria and formulating indicators by considering the program's goals, values and context. Strategic choices consider evaluation paradigms and emphasize the notion of social justice. Based on strategic choices, methodological choices determine the evaluation's course of action regarding the gathering and analysis of information. For example, choice of evaluation model, choice of paradigm, subject selection, chosen techniques to ensure a rigorous process are all considered as being methodological choices. Since strategic choices are used to frame and anchor the methodological choices, they are to be established at the onset of the evaluation process. Doing so, limits the probability of deviating from objectives which in turn augments the validity of claims so as to offer enlightening recommendations. Hurteau's (1991) contribution is significant to the practice in that it expands upon the concept of the Logic of Evaluation by identifying and making the distinction between the two choices facing all evaluators in their daily practice. Furthermore, by placing strategic choices at the onset of the evaluation, it provides an order of execution so as to ensure rigorous practice.

Also, expanding upon Scriven's (1980) concept of logic, Fournier (1995) distinguishes two levels of logic: General Logic and Working Logic. General Logic follows Scriven's four operations and refers to the evaluation's global strategic

process. The Working Logic rests upon the General Logic yet considers the evaluation's context and, as such, renders operational the evaluation process. Each operation found in the General Logic relates to and interacts with the different aspects of the Working Logic as is illustrated in the following figure:

Figure 2.2. The relationship between Working Logic and General Logic
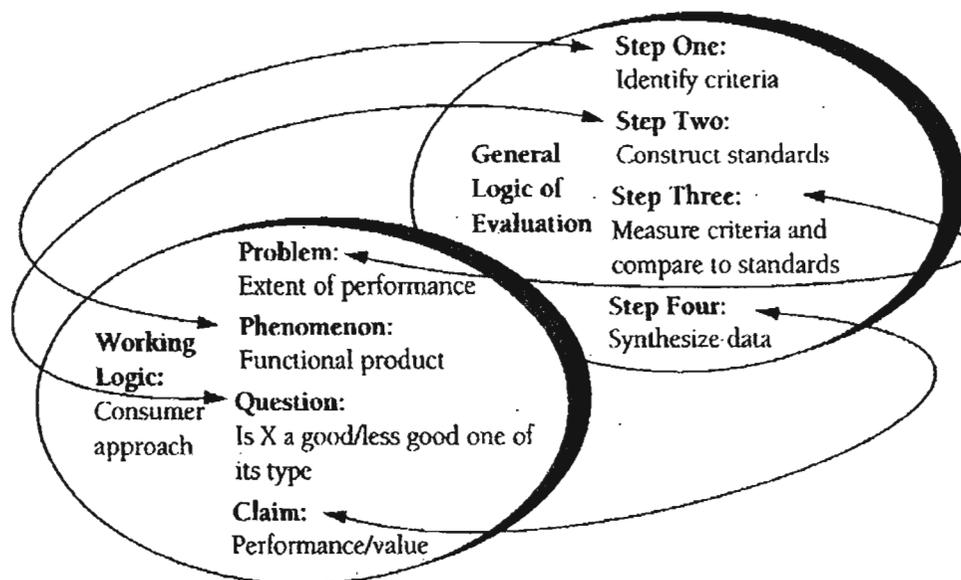
(Fournier, 1995)

Fournier's (1995) contribution to the practice is significant in that the Working Logic introduces parameters which renders operational Scriven's concept of the Logic of Evaluation thus providing a practical aspect to his four basic operations so as to render judgements that are justifiable.

Stake (2004) provides insight into the underlying thought processes involved when approaching an evaluation. He also acknowledges the dual nature of the practice by

identifying critical thinking (standards-based evaluation) and responsive thinking as being two distinct approaches to the evaluation process. Critical thinking refers to evaluations based on standards which focus on rigor, reasoning and logics. It stems from the evaluator's objectivity and follows Scriven's (1980) contribution by emphasizing the need for identifying criteria, establishing standards and measuring variables in order to render valid judgments.

1.2 Elements of the modeling of the program evaluation process

As mentioned at the beginning of this section, all authors who were presented have pursued Scriven's (1980) original reflection so as to, on the one hand, identify the elements necessary in integrating the operations into an functional framework for the practice and, on the other hand, to consider the conceptual and methodological developments that have occurred since Scriven's proposed definition. In their recent study, Hurteau, Lachapelle & Houle (2006) have synthesized the elements of the evaluation process and illustrated them as appearing in figure 2.3 (on the next page). The modeling of the program evaluation process was validated in the first phase of their study (Hurteau & Houle, 2005a).

1.2.1 Element 1: Program to be evaluated

The first element of the modeling refers to establishing the existence of a program by considering the program's context, environment and mode of functioning. In other words, it is within this first element that typical characteristics of a program are to be identified. In order to be considered a program, the following five characteristics must be present: input, plan/organization, activity/service, intermediary factors and results. *Input* refers to all that is necessary in creating a program such as a clientele, a need for the program, objectives, human, financial and material resources, to name a

Figure 2.3. A modeling of the program evaluation process: components and interrelations

Description of the Program to be Evaluated:

Establish if there is a program, understand its components, its theoretical bases and context

↓

Establishing the Evaluation's Dilemma at the Onset of the Evaluation:

Documenting the aspects of the dilemma which motivated the commissioning of the evaluation, identify the client(s) and their concerns

↓

Rendering the Evaluation Operational:

Translated the preoccupations that emerged into evaluation questions, evaluation objectives and evaluation type

↓

Strategic Choices:

Object (s)

Criteria/		Criteria/	
observable variables ◄──► standards		observable variables ◄──► standards	

· · · · ·

Evaluative Claims

Synthesis: Evaluative Conclusions

(Hurteau, Lachapelle & Houle, 2006)

few. The *plan or organization* refers to the set of operations used to organize the program, service or intervention. The *activity* or *service* is that which is offered to the clientele. *Intermediary factors* refer to events that occurred in the implementation phase (i.e. change in personnel, new competitors, change in legislation, etc.). Finally, a program should also show or have anticipated *results* be they short term, medium term or long term. Under certain circumstances, if a substantial gap is found in any of these elements or characteristics, the existence of a program may be questioned.

1.2.2 Element 2: The evaluation's dilemma

The evaluation's dilemma includes documenting factors that motivated the evaluation to be commissioned, identifying the client and targeting the client's preoccupations. This element allows for a better understanding of the client and, as such, will provide the foundation upon which the evaluator's argumentation will be built. If the dilemma is well identified, the judgment rendered will be relevant and persuasive.

1.2.3 Element 3: Rendering the evaluation operational

Rendering the evaluation operational involves translating gathered information into specific evaluation questions and objectives and identifying the type of evaluation that will be conducted. It is worthy to mention that the evaluation objectives and questions are one in the same; differing only in their formulation. They are context specific in that they vary according to the type of argumentation or judgment that is anticipated in the dilemma. They also influence the type of evaluation that will be chosen; needs assessment, assessment of program theory, assessment of program process, impact assessment or efficiency assessment (Rossi, Lipsey and Freeman, 2004). Appendix C provides a more detailed synthesis of this typology.

1.2.4 Element 4: Strategic choices

The fourth element is the strategic choices to be made with regards to the criteria and standards. Originating from Hurteau's (1991) contribution, mentioned in a previous section, this aspect of the operationalization phase involves identifying and defining the object of evaluation, setting the criteria and formulating indicators by considering the program's goals, values and context. These choices create conditions needed to effectively discern and measure factors that influence outcomes.

1.2.5 Element 5: Evaluative claims

The fifth element is the evaluative claims which involve making claims based on the identified criteria and standards. Declarations made should be justified by the predetermined criteria and standards.

1.2.6 Element 6: Synthesis: Evaluative conclusions

The sixth and final element presented in Hurteau, Lachapelle and Houle's (2006) modeling of the program evaluation process is the synthesis which is illustrated by way of the evaluation's conclusions. This element involves deciphering the data gathered through measuring the observed variables which were identified according to the chosen standards. The data is then organized and synthesized into a succinct, coherent conclusion.

These six elements of the modeling of the program evaluation process as proposed by Hurteau, Lachapelle and Houle (2006) concludes this section which synthesizes the elements in order to provide a current modeling of the program evaluation process which is the specific act of evaluating.

This concludes the presentation of our first concept; the modeling of the program evaluation process. We will now present our second concept; the instrumentation elaborated to support the practice.

2. THE INSTRUMENTATION

Hurteau & Houle (2006) proposed a distinction between the modeling of the program evaluation process and the instrumentation elaborated to support it. Although each component of the instrumentation will be presented in the following sub-sections, the program evaluations models/theoretical orientations will be presented in greater detail since they are the main focus of our study.

2.1 Program evaluation models

As will be presented in the following sections, there exist as many evaluation models as there are evaluation theorists. Examining each model individually would certainly prove to be a task nearly impossible to undertake and would risk generating more confusion than providing a new and improved understanding of the field of program evaluation. So as to avoid such eventual confusion, we have chosen to examine evaluation models by that which characterizes them as well as by that which distinguishes them from each other. As such, we refer to Alkin and Christie (2004) who provided a meta structure which categorizes models into three different theoretical orientations. To do so, it is necessary to present the theoretical foundations upon which evaluation models were developed.

2.1.1 Theoretical orientations and their roots: Evaluation models' common attributes

As mentioned, Alkin and Christie (2004) chose to examine evaluation models by searching for common attributes rather than the historically derived relationships between theories which would have been a more traditional approach. In doing so, the authors provide a framework for their analysis that uses the descriptive and prescriptive models whose differences are found within the perspective of the paradigms. In this sense, their presentation is both original and innovative.

Alkin and Christie's (2004) Evaluation Theory Tree serves as a metaphor to better understand the evolution that has occurred in the field of evaluation over the past 40 years thus clarifying the foundations on which countless models were based. Although the authors acknowledge that there may exist different categories, they have chosen to focus their work on three orientations so as to provide a coherent framework within a category system inquiry approach. In order to illustrate each theorist's position, Alkin and Christie (2004) elaborated the Evaluation Theory Tree presented on the next page.

At its base, Alkin and Christie identified Accountability & Control and Social Inquiry as being at the root of program evaluation. In this sense, these two areas serve as motivational factors that emphasize the need for evaluations to be conducted and the value attributed to the use of sound methods needed to effectively complete the process. As such, Accountability & Control and Social Inquiry constitute the trunk of the Evaluation Theory Tree.

Figure 2.4. Evaluation theory tree
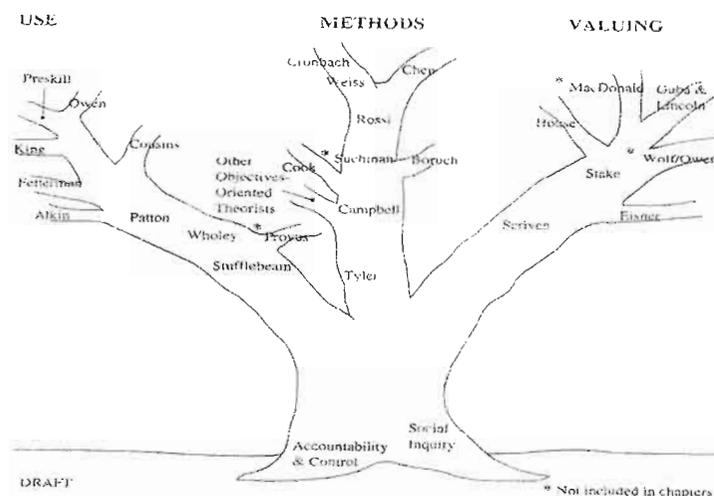
(Alkin and Christie, 2004)

Each category or orientation is represented as a branch on the Evaluation Theory Tree. Using Tyler's (1942) "Eight-Year Study" as a catalyst, the authors identify three theoretical orientations; methods, valuing and use. These categories or orientations constitute the branches of the Evaluation Theory Tree. The methods branch is an extension of the social inquiry trunk and was designated as such "since in its purest form, it deals with obtaining generalizability, or "knowledge construction" as Shadish, Cook, and Leviton (1991) refer to it." (p.12-13). Inspired by Scriven (1967), the valuing branch is home to those theorists who "maintain that placing value on data is perhaps the most essential component of the evaluator's work" (p.13). Theorists who are mainly concerned with the use and the users of evaluation findings are part of the use branch. Alkin's own theoretical orientation may have contributed to the inclusion of the use branch in the Evaluation Theory Tree. Thus, the methods branch consists of theorists whose main concern focuses on the methodology used to obtain information whereas the valuing branch consists of theorists who emphasize the value

of the data obtained. Furthermore, theorists placed on the use branch focus their efforts on how the data obtained will be used. However, Alkin and Christie emphasize that the branches are not mutually exclusive categories. They are based upon the concept of priority and although the placement of a particular theorist may be on one branch, the theorist may be influenced by models on other branches.

Within each branch, theorists were strategically placed in order to emphasize their position in relation to other theorists. By tracing their theoretical influences and considering their published work, Alkin and Christie (2004) placed each theorist within a particular theoretical orientation. Although admittedly difficult at times to categorize, the theorists appearing on the Evaluation Theory Tree have demonstrated, in the authors' opinion, a particular evaluation orientation.

Although Alkin and Christie (2004) have chosen to categorize theorists on one of their three branches, they are aware that their Evaluation Theory Tree is a framework that can be deconstructed to create an entirely new framework. Nonetheless, the orientations presented have been identified by using a category systems approach. This approach involves selecting categories that have been based on a limited set of characteristics and placing elements (in this case theorists) within each category according to a predetermined set of prominent features. Arguably subjective in its approach, category systems allow for a way to identify key characteristics for grouping theories and offer an opportunity for re-examination and other nuances to be made. Nonetheless, the authors chose methods, valuing and use to serve as branches since those were the orientations that aligned with the key characteristics upon which Alkin and Christie (2004) chose to focus.

The following sections will present the two fundamental areas that have been identified by Alkin and Christie as being the root of program evaluation and the three orientations that constitute the branches of the Evaluation Theory Tree.

2.1.2 The root of program evaluation

As mentioned, Alkin & Christie (2004) identified Accountability & Control and Social Inquiry as being fundamental motivational factors which support "the development of the field in different ways" (p.12). For this reason, these two areas or motivational factors constitute the trunk of the Evaluation Theory Tree.

According to the authors, accountability is a process which examines the extent to which a program has established reasonable and appropriate goals, procedures and means to render it accountable. Evaluations are often commissioned for the purposes of providing information in a quest for "being answerable". For example, the Ministry of Education may need to investigate how funding for curriculum reform is being spent. In this case, the educational institutions who received funding from the Ministry of Education would be held accountable or "answerable" to the formal authority that provided the funds (i.e. the Ministry of Education).

Alkin and Christie also consider Social Inquiry as a fundamental area from which evaluation models have been derived. They maintain that Social Inquiry, in its broadest sense, is the study of individual or group behaviour in various social settings using a variety of methods and, in this sense, is composed of a significant social dimension. Social Inquiry also has a strong methodological component in that it "emanates from a concern for employing a systematic and justifiable set of methods" (p.12). Thus, Social Inquiry provides methods in determining how to improve and better programs. Debates surrounding the methodological approaches used in this type of inquiry are ongoing. Nonetheless, most evaluation models find their origin in Social Inquiry.

Alkin and Christie define two general types of evaluation models. First, a prescriptive model offers a "set of rules, prescriptions, and prohibitions and guiding frameworks

that specify what a good or proper evaluation is and how evaluation should be done" (p.5). The second evaluation model type is a descriptive model which "is a set of statements and generalizations which describes, predicts, or explains evaluation activities" (p.5). According to Alkin and Christie, prescriptive models serve as exemplars whereas descriptive models are designed to offer an empirical theory.

The commonality between Accountability & Control and Social Inquiry that places these two areas in the trunk of the Evaluation Theory Tree is that they both provide a foundation on which to either commission evaluations or build theories or models. However, Accountability & Control addresses the need for evaluation whereas Social Inquiry provides the methodology required to adequately measure accountability.

2.1.3 Evaluation orientations: differentiating evaluation models

From Accountability & Control and Social Inquiry, three theoretical orientations were identified by Alkin and Christie; methods, valuing, and use. Each evaluation orientation is illustrated as a branch on the Evaluation Theory Tree. Each branch is either an extension of the social inquiry trunk (i.e. methods and valuing) or of the accountability & control trunk (i.e. use). That is not to say that the branches are representative of a mutually exclusive category system. Alkin (2004) explains the reasoning behind placing theorists on a given branch as follows:

> The distinction between evaluation models based upon these three dimensions is not based on exclusivity, for example, that only one model believes in the use of methodology and others do not. Rather, the category system is based upon relative emphasis within the various models. It might then be possible to ask this question: When evaluators must make concessions, what do they most easily give up and what do they most tenuously defend? (p.8)

Thus, the different branches are admittedly simplistic representations of theorists' main influential factors regarding the choices they make when placed in situations that offer various options.

2.1.3.1 Methods orientation

Finding its roots in the social sciences, the methods orientation stems from Social Inquiry. Evaluation models based on methodology provide theories aimed at increasing the use of appropriate and justifiable methods when studying society, social groups and social life and, in this sense, have strong ties to the social inquiry trunk of Alkin and Christie's (2004) Evaluation Theory Tree. As is the concern of most theorists, those placed on the methods branch of the Evaluation Theory Tree emphasize that scientific research is the result of a well-designed experimental study. However, methodology theorists can be distinguished from other authors in that when they have a choice to make between, for example, scientific rigor and pertinence of the approach they favour scientific rigor whereas "valuing" theorists would favour the latter.

2.1.3.2 Valuing orientation

Alkin and Christie (2004) have identified certain theorists whose work focuses on the making of value judgments. These theorists believe (to varying degrees) that it is the role of the evaluator to provide society with the information needed to make the right choices. Thus, the valuing branch is rooted in the social sciences and is an extension of the Social Inquiry trunk of the Evaluation Theory Tree.

Theorists placed on the valuing branch focus their efforts on finding the true value of that which is being evaluated (i.e. the program or product is either "good" or "bad"). In order to do so, valuing theorists employ a comparative research approach to

evaluation so as to determine appropriate criteria and identify critical competitors. This approach gives way for value judgments to be rendered so as to provide critical information to the general public.

2.1.3.3 Use orientation

Theorists from the use perspective are concerned with the use of findings included in an evaluation once the report is given to those who commissioned the evaluation. The use orientation stems from a need to prove that a program has established reasonable and appropriate goals, that appropriate procedures for attaining those goals have been implemented or that established goals have been achieved in order to improve existing programs or policies or to obtain additional funding. Thus, the use branch is an extension of the Accountability & Control trunk on Alkin and Christie's (2004) Evaluation Theory Tree.

In order to provide information concerning accountability, evaluators must give due consideration to those who are responsible for a program since, in essence, they are accountable for the program's success or failure (Alkin, 2004). However, evaluations do not reflect this notion. Nonetheless, use theorists are well aware of the ambiguous nature surrounding the evaluation of individuals rather than the program itself and, as such, design models that include those responsible for the program in the evaluation process.

In their beginnings, utilization models were oriented, almost exclusively, toward the decision-makers and theorists concentrated their efforts toward key stakeholders who had the power to decide the fate of a program. These decision-makers included administrators, policy makers and directors. In the years to follow, other utilization theories emerged in the literature. Building upon the concept of an inclusive evaluation design, evaluation utilization theorists acknowledged the value of all

stakeholders' active participation in the evaluation process. In doing so, theorists from the use perspective agree that stakeholder participation increases the potential for evaluation findings to be used.

2.1.4 Methods, valuing and use: characteristics and differences

Alkin and Christie (2004) identified three major orientations in the field of evaluation which they chose to illustrate as branches on an Evaluation Theory Tree; methods branch, valuing branch, and use branch. As mentioned in the previous sections, these orientations are not mutually exclusive categories. However, there are fundamental differences between the three orientations that make each category unique allowing for each branch to be separate from the others. Table 2.1 (on the following page) summarizes each orientation's distinctive characteristics.

Although methodology is a concern of all theorists, those placed on the methods branch favour traditional scientific rigor above and beyond all other orientations. These theorists build models that are a reflection of the purity of experimental and quasi-experimental research methods found in the social sciences. Although theorists on other branches may be influenced by methods theorists and employ sound methodology when designing their evaluation models, the methods orientation fundamentally differs from other orientations in that it does not direct its results to a specific individual or group (i.e. decision-maker, stakeholders, consumer) nor does it rely on any given model or method (e.g. identification of critical competitors).

Table 2.1 Distinguishing Characteristics of Each Orientation

Orientations	Main objectives	Main characteristics
Methods	Achieve valid results by creating conditions for appropriate experimental and quasi-experimental designs to effectively discern and measure factors that influence outcomes.	• Scientific research methods and techniques are paramount • Strict chronological sequencing of selected field-relevant methods • Demonstrates the purity of experimental and quasi-experimental research methods • Rooted in the methodological traditions of the social sciences; traditional scientific rigor • Emphasis on well-stated goals, objectives and criteria • Focus on program outcomes
Valuing	Search for the true value of a given object, situation or process so that the consumer can make informed decisions regarding his or her own needs.	• Focus on evaluator's role in the making of value judgments; the program/product is either "good" or "bad" • Relies on the identification of critical competitors and competing alternatives so that value judgments can be made thus providing crucial information to the consumer • Intended audience is the consumer • Stakeholder representation in the evaluation report
Use	The ultimate goal of program evaluation is for the results to be used in order to achieve organizational change so as to best serve the client.	• Evaluation is an interactive process which engages different individuals or group of individuals (i.e. decision-makers, stakeholders, staff members, volunteers) in the evaluation process so as to create a sense of responsibility toward the evaluation thus increasing the potential for use of findings. • Stakeholder participation throughout the evaluation process

Evaluations designed from this perspective focus on the validity of the results by adhering to strict chronological sequencing of selected field-relevant methods. Furthermore, methods theorists thrive for generalizability and create models that can

be useful to all orientations in that they provide, at the very least, "food for thought" when considering the design aspect of any given model.

The valuing orientation distinguishes itself from other perspectives in that it seeks information that facilitates the task of making value judgments. The most obvious distinction between these authors and authors from the use orientation, for example, is that evaluations designed from the valuing perspective aim to inform all consumers on the true value of a program or product so that the consumer or client can make his or her own decision regarding the product or program. Also, the role of the evaluator is to include the different perspectives and communicate alternative definitions without engaging stakeholders any further in the evaluation process. In this sense, the valuing orientation considers stakeholder participation as being punctual.

The use orientation also differs from the methods perspective in that evaluations designed from the use perspective are built upon the notion that the findings must provide information that will be used. Thus, utilization models are built in function of this fundamental issue. As a result, the evaluator becomes much more involved in the evaluation process than would be traditionally acceptable in the methods or valuing orientations. Evaluators from the use orientation actively engage participants and involve themselves in the evaluation process beyond mere observers and recorders so as to increase the potential for use of findings; a price methods oriented theorists would surely not pay for that would mean, to them, a potential for the results to be biased thus jeopardizing the validity of the findings.

2.2 Standards of practice

In 1981, the Joint Committee developed *Standards for Evaluation of Educational Programs, Projects and Materials*. Eventually these standards were used and tested

by various individuals and organisations in their fields. The increased use of the standards by evaluators in different fields and changing contexts for evaluation prompted the Joint Committee to review and revise the standards. A second edition was produced and field tested to "provide a guide for evaluating educational and training programs, projects and materials in a variety of settings" (Joint Committee, 1994).

The *Program Evaluation Standards* are a set of 30 standards divided into four categories: utility, feasibility, propriety and accuracy. Appendix D presents a summary of these standards. The Joint Committee suggests using the standards as a frame of reference to ensure that the evaluator's responsibilities have not been overlooked in the evaluation process. As such, evaluators are encouraged to refer to them regularly through the evaluation process.

2.3 Ethical guidelines

Every major national evaluation society developed its own ethical guidelines which are all very similar in nature. For example, the Canadian Evaluation Society (CES) produced *Guidelines for Ethical Conduct* with three main guidelines (competence, integrity and accountability) and 11 defining statements (Appendix E). The first guideline warrants that evaluators be competent in their provision of services. The second guideline encourages evaluators to act with integrity in their relationships with all stakeholders. The third guideline is intended to ensure that evaluators be accountable for their performance and their product.

Ethical guidelines differ from the standards of practice in that they are used to shed light on ethical dilemmas as they arise. For example, during an interview an evaluator may find that an employee has been treated abusively by the employer. This information has been shared confidentially with the evaluator. The evaluator now

finds himself or herself facing an ethical dilemma. The evaluator must choose between acting upon this information and risk breaching client confidentiality or ignore the information and, as such, risk allowing the abuse to continue. Ethical guidelines can be used as reference to justify the evaluator's position in solving this ethical dilemma. In this case, the evaluator may choose to breach client confidentiality and, in this case, it may be morally right to do so. However, due to other variables that may be present in another case, which may be similar but not identical, the evaluator must assess whether the choice to breach client confidentiality is a viable alternative. In this sense, ethical guidelines are context specific and they are not prescriptive.

Although the *Guidelines for Ethical Conduct* are not accredited, CES members have voted to adopt them. They "are used by many as a touchstone for good practice" (Canadian Evaluation Society).

2.4 Methodological choices

According to Hurteau (1991) methodological choices involve the methods used to gather data, the techniques chosen for data analysis and subject selection and the measures used to ensure that the process was rigorous. Because these choices are not unique to the field of program evaluation, they are considered to be a component of the instrumentation used in the program evaluation process. Also, the modeling of the program evaluation process seeks to identify *what* is being evaluated. The methodological choice clearly address *how* an evaluation will proceed given its context, available resources and all other considerations that vary from one evaluation to the next. Making choices as to how and evaluation will proceed also involves considering *who* will participate in the evaluation process and *when* each phase of the evaluation will take place. According to Hurteau (1991), these choices are to be made at the onset of the evaluation during the planning phase. They are based on the

strategic choices (as seen in section 1.2.4) and specify the procedures required for data collection and information analysis.

3. DISTINCTION WITHIN THE PRACTICE: THE MODELING OF THE PROGRAM EVALUATION PROCESS AND EVALUATION MODELS

Since the main focus of our study is to examine the distinction between the modeling of the evaluation process and evaluation models, it is necessary to offer a framework for this distinction.

Fournier's (1995) introduced parameters that rendered Scriven's concept of the Logic of Evaluation operational thus providing a practical aspect to his four basic operations. She mentions that the Logic of Evaluation, which she refers to as Working Logic, transcends the structures proposed by different approaches. She maintains the following:

> The general logic can be found across various instances of the evaluation inquiry process. For example, the numerous evaluation approaches developed by theorists vary from one another in many details, yet I find that they do share this common logic (Fournier, 1993). [...]What counts as criteria or evidence and how evidence is weighed varies from one approach to another, yet all follow the pattern of evaluative reasoning noted in the four steps. (p.17)

In other words, the modeling of the program evaluation process distinguishes itself from evaluation models in that its elements remain unchanged regardless of the approach taken. They are also interrelated in that although, each approach has its own method of selecting the criteria, establishing standards and synthesizing the information into a final judgement, they nonetheless follow this logical process.

This observation is supported by Alkin and Christie (2004) who categorized evaluation models within a given theoretical perspective (methods, use, valuing).

There exist countless models that propose various methods and components that are adapted to the evaluation's particular objectives or to an evaluator's theoretical orientation. Conversely, the modeling of the program evaluation process offers invariable components that are universal and are the foundation upon which all evaluations rest and as such, transcends all methodological considerations regarding the choice of models. It is a modeling that anchors evaluations' claims by structuring the evaluation process through a systematic logical sequence thus augmenting the validity of results. As Fournier (1995) theorized, the models may provide peripheral parameters on which methodological choices may be made, but they should not influence the fundamental structure of the practice that is necessary in founding valid judgments and claims. It is within this perspective that Hurteau & Houle (2006) propose that the elements of the modeling of the program evaluation process are common to the act of evaluating and, as such, distinguish themselves from evaluation models. As presented in the previous chapter, Figure 2.1 illustrates the distinction between the modeling of the evaluation process and the instrumentation as well as their interrelation.

Figure 2.1: The modeling of the program evaluation process and the instrumentation
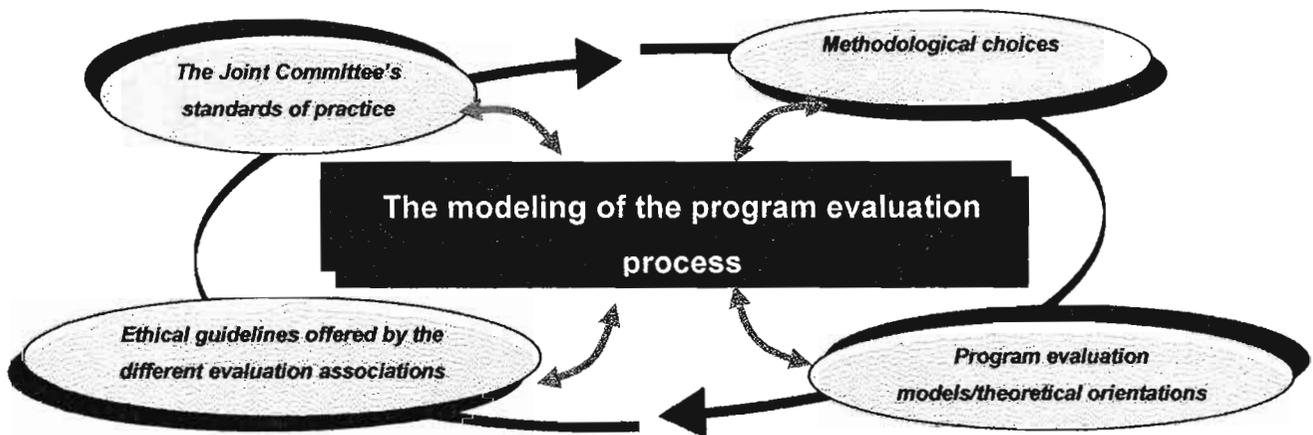
Hurteau & Houle, 2006

In this sense, the modeling of the program evaluation process provides the foundation which constitutes and defines the practice of program evaluation and, as such is a

generic representation of the practice whereas the evaluation models (which are a component of the instrumentation) support the application of these elements within their own given framework and, as such, are interrelated with the evaluation process. However, this reflection has yet to be examined empirically.

4. HYPOTHESIS AND SUB-HYPOTHESIS

As mentioned, the problem raised in the previous chapter refers to the confusion that seems to exist in the literature regarding the distinction between the modeling of the evaluation process and program evaluation models. The purpose of our study seeks to address this problem by considering the modeling which illustrates the act specific to evaluation and evaluation models which are a component of the instrumentation developed to support this act. We chose evaluation models as our starting point because of the apparent confusion that seemingly exists in the literature.

Inspired by Fournier's (1995) and Hurteau and Houle's (2006) reflections which theorized that although evaluation models and approaches may change *how* an evaluation is conducted, they will not change *what* is being evaluated, the following overreaching hypothesis was formulated to guide our study:

Hypothesis: If the elements of the modeling are a generic representation of the program evaluation process, they should be present in the evaluation models.

This overreaching hypothesis seeks to establish the presence of elements that are invariable within the practice which is the act of evaluating. As such, it examines *what* is being evaluated. In theory that which is being evaluated should not vary from one response to the next.

To further orient the hypothesis, a sub hypothesis was formulated.

> Sub-hypothesis: If the modeling of the program evaluation process is indeed generic and if differences are observed in the evaluation models, then these differences should be at the level of the instrumentation.

This sub-hypothesis further orients the overreaching hypothesis in that if the presence of the elements is established, it seeks to address whether the instrumentation used to establish the elements of the modeling varies from one orientation to the next and whether the instrumentation (or components of it) influences the fundamental structure of the practice. In this sense, the sub-hypothesis adds clarity to the results obtained via the hypothesis. In other words, the hypothesis focuses on the modeling of the practice which questions *what* is being evaluated (the evaluand) whereas the sub-hypothesis focuses on the instrumentation which question *how* the evaluand will be evaluated.

CHAPTER III

METHODOLOGY

In order to verify the hypothesis and sub hypothesis presented in the previous chapter, careful consideration must be given to methodological choices so as to augment the validity of the findings. This third chapter will present the research typology and justify the methodological choices. It will state their relevance to the hypothesis and sub-hypothesis and present a research design that will identify the sample as well as the procedures involved in the analysis of data. Finally, due consideration is given to our study's strengths and limitations.

1. THE RESEARCH TYPOLOGY

The present study is influenced by an interpretive epistemological approach in that the knowledge produced is viewed as being intimately related to the context in which it is produced and this knowledge is viewed as being transferable to other contexts (Karsenti and Savoie-Zajc, 2004). However, the extent to which the results are transferable will be discussed in the limits of our study. Positivist influences are also present in that a hypothesis is utilized in the quest for truth.

The present study is descriptive in nature as it seeks to describe the way program evaluation practice (elements of the modeling of the program evaluation process) is rendered operational within the three theoretical orientations and whether differences exist. In the event that differences are identified, the level at which they exist will be presented by considering the instrumentation utilized to establish the elements of the modeling of the program evaluation process.

2. METHODOLOGICAL CHOICES

In order to ensure relevant, credible and valid results, careful consideration has been given to the study's methodological choices. The chosen sample and the type of information required ensure that the content analysis will be relevant. Data collection and analysis have followed a rigorous process which enhances the study's credibility. Standards have been established and applied in order to augment the validity of the results. Each of these aspects is developed in the following sub-sections.

2.1 Type of information required

The overall purpose of this study is to examine a distinction between the modeling of the evaluation process and evaluation models. We will also investigate the nature of discrepancies in the event that it is necessary. As such, the information required will be found in the modeling of the program evaluation process and the components of the instrumentation. Hurteau & Houle (2006) identified four components that constitute the instrumentation; standards of practice, methodological choices, evaluation models and ethical guidelines. Since we can not analyse evaluation models by applying them to themselves and since, as presented in our review of literature, we only apply ethical guidelines if a problem occurs, we will retain two components of the instrumentation if further investigation is necessary; the standards of practice and the methodological choices.

2.2 Sample

A study conducted by Eisenberg, Winters and Alkin (2005) has proven to be a valuable source of data for the current context in that these authors simulated an evaluation scenario of an educational program which they submitted to well established practitioners. These practitioners are recognized by the scientific

community not only as being unequivocally rooted in one of the three previously presented orientations but are also recognized as current leaders for their field work and the contributions they have made through their scientific publications.

The simulated scenario, entitled *The Case: Bunche–Da Vinci Learning Partnership Academy* (Eisenberg, Winters and Alkin, 2005), was published in the scientific journal *New Directions for Evaluation*. It is a case description, set in an education context and based on an actual program. This experimental exercise places the evaluator in a simulated situation that requires an evaluation. The directive was to design an evaluation plan based on the submitted scenario according to that which the respondents (selected evaluators) deemed most appropriate. Four respondents participated in this simulated exercise; two from the methods orientation (Gary T. Henry and Stewart I. Donaldson), one from the valuing orientation (Jennifer C. Greene) and one from the use orientation (Jean A. King). In order to examine an equal number of responses for each orientation, it was necessary to select one representative from each orientation. As such, a choice had to be made in order to retain only one respondent from the methods orientation. So as to situate the reader, authors Christie and Azzam, provide a description of theorists' approaches, in which they state " Henry, through his writings (Mark, Henry, and Julnes, 2000), showed that he would not easily be placed in the methods, value, or use category" (p.16). As such, the three respondents retained for the present study are Stewart I. Donaldson (methods orientation representative), Jennifer C. Greene (valuing orientation representative) and Jean A. King (use orientation representative). Thus, the present study utilizes these three responses to the simulated evaluation scenario as its primary source of data.

Alkin and Christie's (2005) *Unraveling Theorists' Evaluation Reality* is a comparative analysis to respondents' approaches which is utilized as a secondary source of data as it provides information on the evaluation approaches presented.

2.3 Data collection and analysis

Considering the nature of our research, two types of data are required; quantitative data and qualitative data. The overreaching hypothesis, "If the elements of the modeling are a generic representation of the program evaluation process, they should be present in the evaluation models", rather calls for quantitative data since it suggests the need of dichotomous data in order to establish the presence of variables. As such, quantitative data was gathered following the grids presented in Appendix A. These grids are the criteria that served to verify our hypothesis. Each grid represents one of the elements of the modeling of the evaluation process as presented and validated by Hurteau, Lachapelle and Houle (2006). The grids were applied to each of the three theoretical orientations. Though some aspects of the grid (synthesis) could not be fully applied since the responses addressed the planning phase of the evaluation process, valuable data concerning the program evaluation process was nonetheless available in the responses.

So as to verify our sub-hypothesis, "If the modeling of the program evaluation process is indeed generic and if differences are observed in the evaluation models, then these differences should be at the level of the instrumentation", qualitative data was required since it suggests a need for content analysis. In order to verify our sub-hypothesis, two criteria were formulated. The first focuses on elements of the methodological choices. The second focuses the Joint Committee's standards of practice.

More precisely, we examined the methodological choices based on the following criterion:

1. The presence of *who* would establish each element , *how* this would occur and *when* this would occur (intensity of involvement)

The choice to examine the methodological choices based on the three aspects *who*, *how* and *when* was inspired by Alkin and Christie (2005) who, when referring to stakeholder involvement, stated "there are substantial differences between theorists in the choice of stakeholders to be included [*who*], the stages at which they participate [*when*] , and the nature of their involvement [*how*] " (p.118). As such, formulating a first criterion which would consider these three aspects and applying them to the broader area encompassed by the methodological choices has been substantiated.

Potential discrepancies in the responses that may have been due to the standards of practice were examined using the following criterion as a guide:

2. The evaluator's responsibility in managing the *who*, *how* and *when*

This second criterion was examined using Stufflebeam's (1999) Program Evaluation Models Metaevaluation Checklist (Appendix F) since the checklist was elaborated to perform metaevaluations of program evaluation models. Also, it is organized according to the Joint Committee Program Evaluation Standards and it allows for a judgment to be rendered regarding program evaluation models.

The first criterion is based on the results obtained through applying the modeling of the program evaluation process to each individual theoretical orientation. The second criterion uses the Joint Committee's Program Evaluation Standards (Appendix D) as a frame of reference since they refer more specifically to evaluator's responsibility. These criteria were applied to each theoretical orientation allowing for content analysis.

2.4 Establishing standards

As mentioned, the verification of our hypothesis requires a quantitative approach which analyses dichotomous data in order to establish the presence of variables. To do so, we established whether the elements of the modeling are present and associated to the required process. Based on Hurteau, Lachapelle & Houle (2006) a standard rate of occurrence of 80% was retained for the purposes of the present study. This standard was applied both globally (to the three orientations) and holistically (to each individual orientation). According to Stake and Schwandt (2006) "for quantitative and qualitative interpretation alike, we need a qualitative "confidence interval" that counterparts the standard error of measurement" (p.416). As such, we applied the 5% standard margin of error to the retained standard of 80% which established an acceptable rate of occurrence at 75%.

In order to verify our sub-hypothesis, the Joint Committee's Program Evaluation Standards are used as a frame of reference since these standards refer to the evaluator's responsibility (the *who*, the *how* and the *when* aspects). Stufflebeam's (1999) Program Evaluation Models Metaevaluation Checklist (Appendix F), is used as a source of inspiration. It is intended to be used in "performing metaevaluations of program evaluation models and is organized according to the Joint Committee Program Evaluation Standards" (p.1). Also, it provides a standard upon which a judgment regarding program evaluation models can be rendered in that it recommends that "an evaluation be failed if it scores *Poor* on standards P1 Service orientation, A5 Valid Information, A10 Justified Conclusions, **or** A11 Impartial Reporting" (p.1). Finally, the items in the checklist provide a guideline in identifying the level at which discrepancies (if any) occur.

A comparative analysis will examine the methodological choices and, as such, no standard will be offered for this component of the instrumentation

3. THE STUDY'S RIGOR

The study's rigor is found on two levels; theoretical and methodological. On the theoretical level, choices made at the level of the criteria, standards and sample all augment the study's rigor. First, the criteria used to identify the presence of the elements of the modeling (Appendix A) are founded on theory (Scriven, 1980; Hurteau, 1991; and Fournier, 1995). Second, the standards which determined whether the hypothesis and sub-hypothesis were to be confirmed or rejected have been established in the context of existing research. Finally, our sample was chosen based upon the fact that it was published in a scientific journal and, as such, was subject to scrutiny prior to its publication.

On a methodological level, judgments were based on standards that have been documented. That is to say, as mentioned, the standard of 80%, which was related to our hypothesis, was established by Hurteau, Lachapelle & Houle (2006). The Joint Committee's standards of practice, which were used as a frame of reference in verifying our sub-hypothesis, have been adopted by the evaluation community. Furthermore Stufflebeam's (1999) *Program Evaluation Models Metaevaluation Checklist* provides a theoretically supported guideline which helps us to identify discrepancies within the different theoretical orientations and justify our claims. Also, the criterion used to examine the application of the methodological choices was substantiated by Alkin and Christie (2005).

In addition, inter-rater agreement served to increase reliability of the instrument thus leading to an increased reliability of the findings. Inter-rater reliability was measured by evaluating the agreement between 2 people (myself and someone trained to use the instrument). In order to establish the inter-rated agreement the raters applied the grid (Appendix A) to 5 published evaluation reports that were selected at random. Inspired

by Scriven (1995), we attributed a sore in order to weight our criteria. The following point system was used: 2 points were allotted to elements that were identically identified; 1 point was allotted if elements were agreed upon trough discussion and a 0 point was allotted if the raters disagreed on the presence of any given element. Scores were then tabulated and a percentage was calculated. As such, we started data analysis when the inter-rater agreement had been established at 83.4%.

Finally, in order to reduce the risk of error, many techniques were utilized to enhance the rigor of our research. First, the present study follows Miles and Huberman's (1984) and Patton's (2002) suggestions to check the meaning of outliers, generate and assess rival explanations and conclusions, and look for negative evidence and cases. Checking the meaning of outliers is an investigation of exceptions or deviant cases. Discrepancies serve as an alert signal to protect against self-selecting biases. Generating and assessing rival explanations involves looking for data that support alternative explanations. Before generating conclusions, the merits of the "next best" explanation as an alternative was explored. Negative evidence and cases were also considered through actively seeking disconfirming evidence and considering cases and instances that did not fit the patterns and trends that were identified.

Another source of rigor could be had through member checks. This technique "involves taking data, analyses, interpretations, and conclusions back to the participants so that they can judge the accuracy and credibility of the account" (Creswell, p.203). However, the complexity involved with this technique (discussed in the following sub-section) would not be a viable option within this study's context.

4. THE STRENGTHS AND LIMITATIONS OF THE STUDY

Our study's greatest strength is its rigor in that it is anchored in theory at different levels. First, our frame of reference is supported by theory (i.e. the modeling of the program evaluation process, the Joint Committee's standards of practice). Second, as was presented in the previous section, the study's criteria and standards are also supported by theory (Scriven, 1980; Hurteau, 1991; Fournier, 1995; Duval, 2005; Alkin and Christie, 2005). Furthermore, Scriven (1995) inspired us to attribute a score in order to weight our criteria allowing us to calculate a standard so as to confirm or reject our hypothesis on both a global and holistic level. Finally, as mentioned, our sample was published in a scientific journal and, as such, has been scrutinized.

In addition to our rigorous methodology, the study also offers other strengths. First, the selected sample examines one single case from three different theoretical orientations, thus there is only one dependant variable to be examined (the three orientations) since the evaluation content remains the same. This reduces the risk of error. Second, the respondents are designated and recognized by the evaluation community as being representative of one of the three different theoretical orientations. Third, Alkin and Christie (2005) have published a comparative analysis to respondents' approaches entitled *Unraveling Theorists' Evaluation Reality*. This document provides secondary information that is relevant to our inquiry. Finally, the simulated scenario is set in an educational context and, as such, generated results may provide insight to the unique set of challenges facing practitioners when evaluating programs in the field of education.

Among the limitations of the proposed study are those associated with using a simulation or a hypothetical exercise. First, the hypothetical can be somewhat precarious due to a reliance on an incomplete understanding of the context or a lack

of information. These factors may lead respondents to make assumptions and omit certain critical information. Second, respondents might have been unduly influenced by the knowledge that their responses would be published in a scientific journal. It would be important to consider to what extent that knowledge might have influenced their responses. Would their responses be representative of those who have a professional commitment to the field of evaluation? Inquiry into the respondents' intentions and motivations which underlie their published responses would be of interest for future research but is beyond the scope of the proposed study. Third, the sample offers a small number of responses. This decreases the potential for transferability of the results to other contexts. However, it is important to note that the modeling of the program evaluation process has been validated by using a larger sample of published evaluation reports and, as such, some information can be considered as transferable. Also, the information provided only comes from the planning phase. Nonetheless, the responses were identified by Alkin and Christie as being representative of the theoretical orientations and it is within the planning phase that the elements of the modeling are established and, as such, allows us to gather valuable information that is relevant to the proposed inquiry.

A final limitation of the study was introduced in the previous section. Ideally, member checks would be of value in increasing the study's credibility. However, this technique would require, at the very least, a telephone interview with the respondents' in order to familiarize them with the instruments used to analyze the data. In order to appreciate the complexity of such a request, we must remember that these participants were selected, in part, because of their relentless contributions the field and, as such, they have limited availability.

CHAPTER IV

RESULTS

As established in the methodology, globally, the present chapter will present a content analysis, it will treat and analyse the data and it will apply the standard. Also mentioned in the methodology, the results presented in this section are based on information found in three responses to a simulated hypothetical scenario. The three responses have been elaborated from the different theoretical orientations; valuing, methods and use (Alkin and Christie, 2004). They have been cross examined using the following elements of the modeling of the program evaluation process (Hurteau, Lachapelle & Houle, 2006):

1. Description of the Program
2. Establishing the Evaluation's Dilemma at the Onset of the Evaluation
3. Rendering the Evaluation Operational
4. Strategic Choices
5. Evaluative claims
6. Synthesis: Evaluation's Conclusions

As previously presented, the following hypothesis was formulated to guide the first part of the study:

Hypothesis: If the elements of the modeling of the program evaluation process are a generic representation of the program evaluation process, they should be present in the evaluation models.

As mentioned, this first hypothesis seeks to establish the presence of elements that are invariable within the practice which is the act of evaluating. As such, it examines *what* is being evaluated. In theory that which is being evaluated should not vary from one response to the next.

Our sub-hypothesis; If the modeling of the program evaluation process is indeed generic and if differences are observed in the evaluation models, then these differences should be at the level of the instrumentation; further orients the overreaching hypothesis in that once the presence of the elements has been established, the sub-hypothesis seeks to address whether the instrumentation used to establish the elements of the modeling varies from one orientation to the next and whether this instrumentation influences the fundamental structure of the practice. In other words, the hypothesis focuses on the modeling of the process which is hypothesized as being generic and should be invariable from one orientation to the next since it questions *what* is being evaluated (which is the evaluand). On the other hand, the sub hypothesis focuses on the three orientations and seeks *how* the evaluand will be evaluated and, as such, could plausibly change from one orientation to the next.

1. HYPOTHESIS : ESTABLISHING THE PRESENCE OF THE ELEMENTS OF THE MODELING OF THE PROGRAM EVALUATION PROCESS IN THE EVALUATION MODELS

So as to establish the presence of the elements of the modeling in each theoretical orientation, a cross analysis of the presence of the elements of the modeling for each orientation is based on the results from the analysis grids (Appendix A) which have generated the findings concerning the first hypothesis.

Table 4.1 (on the following page) presents a global view of the compilation of the data gathered. The fundamental elements of the modeling of the program evaluation process were recorded as being either present, present in part or absent for each response. The results to each element of the modeling are subsequently presented and developed.

Table 4.1. Global view of the positions held by the three theoretical orientations

Theoretical orientation	Elements of the program to be evaluated are mentioned	Elements of the dilemma regarding client's expectations are mentioned — YES	— In part	— NO	Type of evaluation is mentioned — YES	— NO	Evaluation Questions/Objectives — Present	— In line (w/Rossi, Lipsey & Freeman)	— NO	Quest./Obj./eval. Type consider dilemma — YES In line	— NO	Criteria — Present	— weighted	— NO	Standards — YES	— In part	— NO	Intended Evaluative declarations — Yes	— NO	Intended Declarations founded on — Criteria	— Std.
Valuing	X		X		X		X	X		X		X implicit before emitting the judgment	In part (id'ed eval. priorities)		X (explicitly before emitting the judgment)			X		X (justified implicitly before judgment	X (explicitly justified before judgment)
Methods	X		X		X		X	X		X		X (in line with dilemma before emitting judgment	X		X (in line with dilemma before emitting judgment			Intent is to present results so that stakeholders can make their own judgment		X (explicitly justified before judgment)	X (explicitly justified before judgment)
Use	X		X		X				X		X			X		X (before analysis of results)			No mention of providing a judgment	No intention of providing an argument	

*Although justification was provided for the criteria and the comparison standard, it was not with the intent to provide an argument upon which the judgment would be based, since there is no intention, in this orientation a judgment on the program's worth or merit.

1.1 Element 1 of the modeling of the program evaluation process: Description of the program to be evaluated

Four elements were observed: 1) the clientele; 2) the program's objectives; 3) the intervention's characteristics; and 4) the program's theoretical framework. According to the results obtained, it is established that all four elements were explicitly present in each response. As presented in Table 4.2 the rate of occurrence for this first element is 100% in all three orientations. It can thus be concluded that this first element is present in all three theoretical orientations.

Table 4.2. Results of the existence of a program to be evaluated

Theoretical orientations	The clientele is mentioned		The program's objectives are mentioned		The intervention's characteristics are mentioned		The program's theoretical framework is mentioned	
	Yes	No	Yes	No	Yes	No	Yes	No
Valuing	X		X		X		X	
Methods	X		X		X		X	
Use	X		X		X		X	
Rate of occurrence	3/3		3/3		3/3		3/3	

1.2 Element 2 of the modeling of the program evaluation process: The evaluation's dilemma

This second element of the modeling seeks to establish whether the needs of the client have been considered at the onset of the evaluation. Three indicators were considered for observation; 1) mention of the evaluation's triggering factor; 2)

whether the client expects the evaluator to attribute an order, a score or a mark to the object under evaluation (Type of judgment Scriven,1995); and 3) mentioned whether the client expects the evaluator to support the judgment by basing the evaluation on the judgment of an expert, the causal link between two variables, the perception of the clientele or of the stakeholders, the inherent properties of the intervention, or the needs of the clientele (criteria sources according to Fournier, 1995). These results are presented in Table 4.3 as follows:

Table 4.3. Results of the presence of the evaluation's dilemma in each orientation

Theoretical orientations	Presence of the evaluation's triggering factor		Client's expectation as to the type of judgment rendered (Scriven, 1995)		Client's expectation as to the source of criteria on which the judgment will be based (Fournier, 1995)	
	Yes	No	Yes	No	Yes	No
Valuing		X		X	X	
Methods	X			X	X	
Use	X			X	X	
Rate of occurrence	2/3		0/3		3/3	

Similar to the first element, the results obtained for this second element on all aspects at the exception of the descriptor which refers to the client's expectation regarding the judgment to be rendered. Both the methods and the use orientations specifically mentioned the evaluation's triggering factor whereas the valuing orientation overlooked this aspect of the second element of the modeling. Nonetheless, the two remaining aspects yielded identical results. None of the three orientations made

mention regarding the client's expectation as to the type of judgment to be rendered (Scriven, 1995) yet all three orientations provided information regarding the source of criteria that could be linked to Fournier (1995).

Thus, findings for this second element show that two of the three descriptors related to the evaluation's dilemma are present in the methods and use orientations and two of the three descriptors are absent from the response provided by the valuing respondent. As such, globally it can be concluded that since two of the three indicators of the element can generally be found in the responses, this second element is present to some degree in each theoretical orientation.

1.3 Element 3 of the modeling of the program evaluation process: Rendering the evaluation operational

This third element of the modeling seeks to establish whether the evaluation was rendered operational. The following indicators were considered for observation; mention of the type (s) of evaluation, specification of the evaluation question(s) and/or objective(s) in line with the context, whether these indicators are linked to Rossi, Lipsey and Freemean's (2004) typology (Appendix C) and whether these elements consider the elements found in the dilemma. Table 4.4 (on the following page) presents the results.

Results show that all three orientations explicitly identified the type of evaluation. Two of the three orientations (methods and valuing) were identified as providing explicit information regarding the presence of all fundamental elements of the operationalization phase.

In both these cases, fundamental elements that render the evaluation operational definitively considered at least one of the two aspects of the dilemma.

Table 4.4. Results regarding the fundamental elements which render the evaluation operational

Theoretical orientations	Are the fundamental elements of the operationalization phase addressed?				Are these elements linked to Rossi, Lipsey and Freemean's (2004) typology?			Do these elements consider the elements found in the dilemma?				
	The type(s) of evaluation is/are specified		The question(s) and/or the objective(s) (which specify the type of evaluation according to the context) is/are mentioned		Yes	No		These elements are in line with the triggering factor		These elements are in line with Scriven **or** Fournier's criteria		
										Yes	No	
	Yes, explicitly	No, it must be deduced or there is no mention of it	Yes, explicitly	No, it must be deduced or there is no mention of it		The typology is not respected	Lack of info.	Yes	No (lack of info.)		They are not considered	Lack of info.
Valuing	X		X		X			X				X
Methods	X		X		X			X		X		
Use	X			X			X	X				X
Rate of occurrence	3/3		2/3		2/3			2/3		1/3		

Results regarding the third orientation (use) were found to be inconclusive due to a lack of information. Thus, it can be concluded that this third element is present in two of the three theoretical orientations.

1.4 Element 4 of the modeling of the program evaluation process: Strategic choices

This fourth element of the modeling requires of the evaluator to choose specific criteria and standards in line with the evaluation's dilemma prior to the analysis of results. The following indicators were considered for observation; mention of the criteria (elements of the object that will be evaluated), mention of the standards (factors that will determine whether a favorable or unfavorable judgment will be rendered), mention of attributing a relative weight to the criteria, whether these elements consider the elements of the dilemma and whether the criteria and standards presented before the analysis of results. Table 4.5 (on the following page) presents the results.

Results show that all three orientations offered information regarding the evaluation's strategic choices as part of their response. More specifically, the three orientations presented the strategic choices' fundamental elements either entirely or in part. As well, the three responses provided information that either definitely or at least in part links the strategic choices to the evaluation's dilemma. Also two of the three orientations (valuing and methods) presented all the criteria and standards prior to the analysis of results and the third orientation (use) presented some before with the understanding that others would emerge as the evaluation process evolved.

Thus it can be concluded that all indicators related to the strategic choices were found to be present to varying degrees in all three responses.

Table 4.5. Results regarding the evaluation's strategic choices

Theoretical Orientations	Are the strategic choices' fundamental elements presented?									Do these elements consider the elements of the dilemma?			Are the criteria and standards presented before the analysis of results?					
	The criteria (elements of the object that will be evaluated) have been considered			The standards (factors that will determine whether a favorable or unfavorable judgment will be rendered) have been considered			Consideration has been given to attributing a relative weight to the criteria						Criteria			Standards		
	Yes	In part	No	Yes	In part	No	yes	In part	No	Yes	In part	No	Yes	Some or all emerge along the way	No	Yes	Some or all emerge along the way	No
Valuing	X			X				X		X			X			X		
Methods	X			X				X		X			X			X		
Use	X				X			X		X				X			X	
*Rate of occurrence	3/3			2/3			0/3			2/3			2/3			2/3		

*For the purposes of this section only "yes" responses were considered as part of the rate of occurrence. However, "in part" responses will be considered in subsequent analyses.

1.5 Element 5 of the modeling of the program evaluation process: Evaluative Claims

Although six elements are included in modeling of the program evaluation process, we were unable to retrieve information regarding the sixth element of the modeling (synthesis: evaluation conclusions) since the responses focus on the planning phase of the evaluation. Nonetheless we were able to obtain valuable information regarding the fourth element of the modeling. Therefore the final element observed in this study is the evaluative claims.

Results (presented in Table 4.6) show that, at least in theory, two of the three orientations (valuing and methods) intended on providing a value judgment of the object under evaluation and that these two orientations justified either implicitly or explicitly their choice of criteria and standards upon which their judgment will be based. Whether the respondents would actually follow through on their intentions is unknown since we are dealing in the hypothetical. On the other hand, if these intentions were made known there is no reason to believe they would not occur in practice.

Table 4.6. Results regarding the evaluation's evaluative claims

Theoretical orientations	Intent to provide a judgment regarding the program's merit or worth		Justification is provided for the criteria		Justification is provided for the standard	
	Yes	No	Yes	No	Yes	No
Valuing	X		X		X	
Methods	X		X		X	
Use		X	X		X	
Rate of occurrence	2/3		3/3		3/3	

1.6 Synthesis of the results pertaining to the hypothesis

Table 4.7 (on the next page) summarizes the results pertaining to the rate of occurrence of the elements of the modeling (presented in table 4.1) in each theoretical orientation. As presented in Table 4.1, each element of the modeling has a certain number of identified indicators that were used to observe and establish the presence of the elements in each theoretical orientation. The average rate of occurrence for each element was calculated based on the following point system:

- 2 points were allotted to the indicator if it was found to be present in the response
- 1 point was allotted to the indicator if it was found to be present in part
- A score of 0 was allotted if the indicator was not present in the response

Results presented in Table 4.7 will be developed in the following sub-section.

Table 4.7 Summary of results: Identified rate of occurrence for elements of the modeling
of the program evaluation process

Elements	Valuing Orientation	Methods Orientation	Use Orientation	Rate of Occurrence for each item
Program to be evaluated:				
1. Clientele is mentioned	2	2	2	6/6
2. Program objectives are identified	2	2	2	6/6
3. Intervention's characteristics are mentioned	2	2	2	6/6
4. Program's theoretical framework	2	2	2	6/6
The evaluation's dilemma:				
1. The evaluation's triggering factor is mentioned	0	2	2	4/6
2. The report specifies whether the client wants the evaluator to attribute an order, a mark, a score or to assign a class to the object being evaluated (Types of judgment according to Scriven, 1995)	0	0	0	0/6
3. It is mentioned whether the client expects the evaluator to support the judgment by basing the evaluation on the judgment of an expert, the causal link between two variables, the perception of the clientele or of the stakeholders, the inherent properties of the intervention, or the needs of the clientele (Criteria sources according to Fournier, 1995)	2	2	2	6/6
Rendering the evaluation operational:				
1. The type(s) of evaluation is/are specified	2	2	2	6/6
2. The question(s) and/or the objective(s) (which specify the type of evaluation according to the evaluation context) is/are mentioned	2	2	0	4/6
3. The type(s) of evaluation and the questions/objectives are linked to Rossi, Lipsey, Freeman's (2004) typology	2	2	0	4/6
4. These elements are in-line with the triggering factor	2	2	0	4/6
5. These elements are inline with Scriven's or Fournier's criteria	0	2	0	2/6
The evaluation's strategic choices:				
1. The criteria (elements of the object that will be evaluated) have been considered	2	2	2	6/6
2. The standards (factors that will determine whether a favorable or unfavorable judgment will be rendered) have been considered	2	2	1	5/6
3. Consideration has been given to attributing a relative weight to the criteria	1	1	1	3/6
4. These elements consider the elements of the dilemma	1	2	2	5/6
5. The criteria and standards are presented before the analysis of results	2	2	1	5/6
Evaluative Claims:				
1. The intent is to provide a judgment regarding the program's merit or worth	2	2	0	4/6
2. Justification is provided for the criteria	2	2	2	6/6
3. Justification is provided for the standards	2	2	2	6/6
Total score for each orientation	**32/40**	**37/40**	**25/40**	**94/120**

1.7 Conclusion of the hypothesis

In order to verify the hypothesis we announced a standard of 80% that we can look at globally and holistically. To be able to apply the standard we translated the information presented in Table 4.7and further reduced it thus enabling us to apply the standard of 80%. In order to be able to apply the standard we needed to translate scores into a percentage. To do so we used the following formula:

$$\frac{\text{Score of observed occurrence of the element}}{\text{Total possible score for the element}} = \frac{X}{100}$$

As mentioned in Chapter 3, the minimum rate of occurrence was established at 80% when we apply the 5% margin of error of we can tolerate a minimum standard rate of occurrence of 75%.

To identify the level at which the discrepancies occur, additional information was also considered. As such, scores for each element of the modeling were tabulated and transformed into a percentage indicating the average rate of occurrence of the element in each theoretical orientation. An overall average rate of occurrence for each element of the modeling was then established. The results are presented in Table 4.8 (on the next page).

The elements of the modeling of the process were observed in all three orientations to varying degrees. The methods orientation obtained an average rate of occurrence of 92.5%, the valuing orientation obtained an average rate of occurrence of 77.5% and the use orientation obtained an average rate of occurrence of 62.5%. Overall, the elements of the modeling were found to be present 78% of the time.

Table 4.8. Occurrence of the elements of the modeling in each theoretical orientation

Elements of the modeling	Average rate of occurrence for each element in each theoretical orientation			
	Valuing Orientation	Methods Orientation	Use Orientation	Overall Average rate of occurrence of each element (rounded off)
Program	100%	100%	100%	100%
Dilemma	33.3%	66.7%	66.7%	56%
Rendering the evaluation operational	80%	100%	20%	67%
Strategic Choices	80%	90%	70%	77%
Evaluative claims	100%	100%	66.7%	89%
TOTAL of each orientation	80 %	92.5 %	62.5 %	**N** **78%**

The confirmation or rejection of the study's following over reaching hypothesis; If the elements of the modeling are a generic representation of the program evaluation process, they should be present in the evaluation models; rests upon our announced standard. Since the standard was set at a minimum rate of occurrence of 80% and considering the 5% margin of error, the threshold is established at a 75% rate of

occurrence. That is to say, that the elements of the modeling of the process must obtain a total rate of occurrence of 75% or higher.

As mentioned, two types of standards were applied in the current study; the global standard (the application of the standard to the global result) and the holistic standard (the application of the standard to each individual result). Since, for the purposes of the current study, we have decided to apply both types of standards, the results presented in Table 4.8 show that the overreaching hypothesis is confirmed if we retain the global standard which situates the global score at a 78% rate of occurrence. However, the hypothesis is rejected if we apply the holistic standard since the rate of occurrence for the use orientation is only at 63%. Since we are considering both types of standards we must conclude that our hypothesis is confirmed in part. That is to say, considering our global rate of occurrence, there is no significant statistical variations. The variations occur at the level of the content and, as such, the hypothesis is partly accepted. However, it is important to specify that this judgment is based on written responses since we were unable to interview the respondents. As a result, findings presented in our research may have been different had we conducted an interview to probe the respondents' true intents.

Finally, another interesting finding emerged from the data. Prior to signing the contract with the client, all three orientations identified the negotiation process as being a key factor in a successful evaluator-client relationship. The implications of this finding will be discussed in the next chapter.

To conclude, since a discrepancy has been identified at the level of the holistic standard, it is necessary to delve deeper into our investigation in order to examine whether the discrepancy is due to the instrumentation used to identify the evaluand or whether it is due to the modeling not being generic. The following section will apply the two retained components of the instrumentation (the methodological choices and

the standards of practice) to the three theoretical orientations so as to establish the level at which the differences exist.

2. SUB-HYPOTHESIS: IF THE MODELING IS INDEED GENERIC AND IF DIFFERENCES ARE OBSERVED IN THE EVALUATION MODELS, THEN THESE DIFFERENCES SHOULD BE AT THE LEVEL OF THE INSTRUMENTATION.

As mentioned, this second hypothesis is used to further orient the first hypothesis and as such is considered a sub-hypothesis of the first overreaching hypothesis. This sub-hypothesis seeks to address whether the instrumentation used to establish the elements of the modeling of the program evaluation process varies from one orientation to the next and whether this instrumentation influences the fundamental structure of the practice. In other words, the sub-hypothesis is used to delve deeper into the hypothesis, a necessary step since the results linked to the hypothesis do not allow us to clearly identify whether the discrepancies reside in the fundamental structure of the practice or whether they are found in the instrumentation elaborated to support the practice.

As mentioned in the theoretical framework, the instrumentation elaborated to support the practice includes evaluation models, ethical guidelines, methodological choices and standards of practice (Hurteau & Houle, 2006). The evaluation models are intentionally excluded from this investigation since they constitute the object of the present study. Also, as previously mentioned, ethical considerations are difficult to examine in the present context. As such, we chose to retain the 2 criteria to guide our study regarding the sub-hypothesis. The first focuses on elements of the methodological choices. The second focuses the Joint Committee's standards of practice.

More precisely, we examined the methodological choices based on the following criterion:

1. The presence of *who* would establish each element , *how* this would occur and *when* this would occur (intensity of involvement)

As mentioned, the choice to examine the methodological choices based on the three aspects *who*, *how* and *when* was inspired by Alkin and Christie (2005) who, when referring to stakeholder involvement, stated "there are substantial differences between theorists in the choice of stakeholders to be included [*who*], the stages at which they participate [*when*] , and the nature of their involvement [*how*] " (p.118). As such, formulating a first criterion which would consider these three aspects and applying them to the broader area encompassed by the methodological choices has been substantiated.

Potential discrepancies in the responses that may have been due to the standards of practice were examined using the following criterion as a guide:

2. The evaluator's responsibility in managing the *who*, *how* and *when*

This second criterion was examined using Stufflebeam's (1999) *Program Evaluation Models Metaevaluation Checklist* (Appendix F) since the checklist was elaborated to perform metaevaluations of program evaluation models. Also, it is organized according to the Joint Committee Program Evaluation Standards and it allows for a judgment to be rendered regarding program evaluation models.

Each criterion will now be developed in the following sections.

2.1 Methodological Choices: Modality: Content Analysis

So as to establish whether discrepancies can be identified within the methodological choices, responses were analysed based on our first criterion (the presence of *who* would establish each element, *how* this would occur and *when* this would occur (intensity of involvement). The table presented in Appendix B summarizes the comparative analysis of the results. Each aspect will be discussed in the following sub-sections.

2.1.1 Aspect 1: Who is involved in the evaluation process?

Similarities

Findings show that all three orientations have expressed a need for an evaluation team comprised of two or more evaluators who share a similar sociocultural background with the community members involved in the program. Team members also possess expertise in the field and knowledge of the context. The input of an expert reviewer was also identified as being valuable in all three orientations and the three respondents viewed the evaluator as a driving force in the evaluation process. As well, all three orientations considered stakeholder participation as a key factor in the evaluation.

Distinctions

Although the methods and the valuing orientations organize their stakeholder groups in a very similar manner (parents, teachers, staff members and developers of the Da Vinci Learning Program), the valuing orientation specifies the importance of including the perspectives of "the more marginalized people in the context" (p.32) such as discouraged teachers and transient families. On the other hand, the use orientation proposes an evaluation group comprised of approximately 25 members who are selected as representatives of the various stakeholder groups. Since the intent

is to create an evaluation infrastructure within the school, the use orientation also requires an evaluation advisory committee (EAC) whose members are carefully selected. The EAC would be comprised of the school principal as well as three teachers who possess a positive attitude toward evaluation in the hopes that they would be easily trained to become internal evaluators and advocates.

Conclusion

Thus, it can be concluded that all three orientations generally agree to involve various groups in the evaluation process. However, the selection criteria for each group vary from one orientation to the next particularly in the case of the use orientation.

2.1.2 Aspect 2: How will the information be gathered and interpreted?

Similarities

The valuing and the methods orientations proceed in a very similar manner. They both express using interviews and discussion groups with the various stakeholder groups. They both describe the process as an interactive or collaborative effort in which stakeholders express their needs and provide information which is interpreted by the evaluation team and serves to formulate evaluation questions and establish objectives and criteria. Interviews and group discussions serve a similar purpose in these two orientations. "The evaluator also contributes to the discussion her or his expertise or perspectives, including ideas from relevant literature"(p.31). Also, team members "assess plausibility of stakeholders' program theory or theories" (p. 74). In this sense, the stakeholders and the evaluators and their team work collaboratively to provide relevant data upon which a judgment on the quality of the program can be rendered.

Distinctions

The use orientation employs a very different approach in that the evaluator trains a few selected EAC members to lead efforts that are considered, in the valuing and methods orientations, the responsibility of the evaluator (and team). It is this sense the intensity of stakeholder involvement is manifested. For example, King, the respondent for the use orientation, describes an activity which involves making sense of test scores as follows:

> I would therefore propose that one or two members of the EAC agree to lead a separate committee that would be charged with studying the school's test scores—both company and state—for the past several years and interpreting them with a view to action. We would access someone (from the evaluation team, the district office, a local university, or research shop) with a good understanding of test interpretation and, ideally, the ability to work with the data to answer targeted questions the group might raise.(pp.94,5)

Evidently, the use orientation perceives the evaluation team members as being evaluation trainers or coaches rather than evaluators commissioned to evaluate the program at hand. In this sense, the use orientation differs from the two other orientations in a very fundamental way; the sharing of the evaluation responsibilities. Both the valuing and the methods orientations consider data interpretation as being the responsibility of the evaluator (and team) whereas the use orientation relinquishes (at least to some degree) this responsibility to primary intended users.

Conclusion

To conclude, both the valuing and the methods orientations have taken similar approaches as to how they propose to gather information. In both orientations, stakeholders serve as content or object "experts". Although the use orientation also uses stakeholders to formulate evaluation questions and criteria, establish evaluation priorities and outcomes it views stakeholder participation in a much broader sense in

that the it goes beyond and asks of the primary intended users to collect and interpret data and, as such, members of this group become members of the evaluation staff. In this sense the intensity of stakeholder involvement differs. The following section will consider its impact on the *when*.

2.1.3 Aspect 3: When will the information be gathered (intensity of stakeholder involvement)?

Similarities

All three orientations foresee a schedule in which information gathering events are planned to take place. The valuing and the methods orientations plan punctual events to consult stakeholders and different stakeholder groups at specific times throughout the evaluation process. Stakeholders contribute specific information and shed light on the evaluand.

Differences

The intensity of stakeholder involvement differs significantly in the use orientation's response as compared to the other two responses in that stakeholder participation is continuous throughout the evaluation process. In this sense, stakeholders become active participants in the evaluation process. This creates another fundamental difference between the use orientation and the two other orientations; the perceived purpose that underlies stakeholder participation. As mentioned, the valuing and the methods orientations view stakeholders as being content or object "experts" and as such they are consulted at very specific times during the evaluation. However, the response provided by the use orientation presented a small team of teachers (primary intended users) who would be trained to become "internal evaluators" in sorts. Such and endeavour requires frequent meetings. As well, the use orientation proposed monthly meetings with a team whose members include representatives of various stakeholder groups.

Conclusion

Although all three orientations have set aside time to integrate stakeholder participation in their proposed evaluation, the frequency of participation differs in the use orientation. Stakeholders who participate in the evaluation proposed by the use orientation are required to do so more frequently then those who participate in the evaluations proposed by the methods and valuing orientations.

2.1.4 Methodological Choices: Conclusion

To investigate possible difference in the methodological choices, observations were based on the following criterion:

1. The presence of *who* would establish each element , *how* this would occur and *when* this would occur (intensity of involvement)

Our findings have identified the presence of *who* would establish each element, *how* this would occur and *when* this would occur in all three orientations. Results show similarities between the valuing and the methods orientations with regards to each aspect of the criterion. Differences were found in the information provided by the response from the use orientation with regards to the frequency and intensity of stakeholder participation.

This finding is inline with Alkin and Christie (2005) who compared the valuing and use orientations' view on stakeholder participation and stated: "In our view, there was a difference in intensity in the engagement of stakeholders" (p.117). Based on our results we can conclude that indeed the use orientation requires of its participants a greater involvement in the evaluation process than do the methods and the valuing orientations.

Another fundamental difference found via the second aspect of this first criterion (*how* will each element be established). The use orientation had a very different view of the evaluator's role in the evaluation process as compared to the methods and valuing orientations. As mentioned, the use orientation presents the evaluator as being an evaluation trainer or coach whereas the valuing and the methods orientation view the evaluator as being commissioned to gather, analyse and interpret data so as to render a judgment on the program's worth or merit.

Does this differing view of the evaluator's role or responsibility toward the evaluation influence the quality of the evaluation? The following section delves deeper into this aspect by using the Joint Committee Program Evaluation Standards as guidance.

2.2 Evaluator's responsibilities according to the Joint Committee's standards of practice

As mentioned in the Review of literature, the Joint Committee on Program Evaluation produced standards of practice to help guide evaluators through the program evaluation process (the how). Evaluators are encouraged to refer to them and integrate them in their practice. They focus on the evaluator's responsibility toward an evaluation. Violation of any of these standards should alert the evaluator to reconsider certain aspects of the proposed evaluation. As such, they are part of the instrumentation elaborated in support of the practice and constitute the second component of the instrumentation retained for investigation. To guide this portion of our study, the second criterion was presented as follows:

2. The evaluator's responsibility in managing the *who*, *how* and *when*

Analysis of this second criterion followed Stufflebeam's (1999) *Program Evaluation Models Metaevaluation Checklist* (Appendix F). We chose the checklist because it

was elaborated to perform metaevaluations of program evaluation models. Also, it is organized according to the Joint Committee Program Evaluation Standards and it allows for a judgment to be rendered regarding program evaluation models. As mentioned, Stufflebeam (1999) recommends that "an evaluation model be failed if it scores Poor on standards P1 Service orientation, A5 Valid Information, A10 Justified Conclusions, or A11 Impartial Reporting" (p.1). As such, those four standards were the focus of analysis which was conducted by following Stufflebeam's directives. That is to say, each indicator formulated by Stufflebeam (1999) was applied to each response then scored according to Stufflebeam's (1999) scale as follows:

9-10 (X) Excellent

7-8 (X) Very Good

5-6 (X) Good

3-4 (X) Fair

0-2 (X) Poor

Some of the indicators found in the checklist are subjective in nature and others are based on intent. To truly assess whether there exists a common understanding of the indicators among the respondents, an interview with each respondent would have been necessary. Since this was not feasible for our current study (as mentioned in the limits of our study), the results obtained were based on the clearly stated information presented in each response.

The results for each of the four crucial standards are presented in the tables found in Appendix G. The following sub-sections will present the content analysis of the results obtained. As reference, a definition of each standard is provided prior to presenting the findings. These definitions were taken verbatim from the Joint Committee's Program Evaluation Standards (Appendix D).

2.2.1 Propriety Standards

The propriety standards are intended to ensure that an evaluation will be conducted legally, ethically, and with due regard for the welfare of those involved in the evaluation, as well as those affected by its results.

P1 Service Orientation--Evaluations should be designed to assist organizations to address and effectively serve the needs of the full range of targeted participants.

The results obtained regarding the first crucial standard (P1 Service Orientation) identified by Stufflebeam (1999) show that the model proposed by the methods orientation received an excellent score (9/10), the model proposed by the valuing orientation was also excellent (9/10) and the model proposed by the use orientation was good (5/10).

Similarities

As shown in the first table in Appendix G, all three orientations obtained identical results regarding 6 out of the 10 indicators. Furthermore, the valuing and methods orientations obtained identical results regarding all 10 indicators of this standard.

Differences

One oversight of the use orientation which may have influenced the results is attributed to the fact that the model it proposed was not necessarily in line with the customer's needs. It required a great deal of participation from primary intended users and stakeholders. The context in which the primary intended users (teaching personnel) were working made them feel overwhelmed and, as a result, the school had a large turnover in personnel. Requiring more time from them may not have been feasible in this context.

Furthermore, the customer needs were quite clearly expressed in the scenario "He [superintended Chase] wanted to support her [principal García] in her efforts, and to do so she would need evaluation help, not only to monitor the impact of her changes but also as a reliable source of evaluative data on the impact of the program or parts of it." (p.13). However, the response from the use orientation admitted "Planning for evaluation capacity building differs from planning for an evaluation. Rather than developing an evaluation design in a traditional sense, I have presented a list of activities in this section that would foster a culture of evaluation [...]" (p. 96).

Conclusion

The main focus of the model proposed by the use orientation is to foster a culture of evaluation rather than to focus on the needs expressed by the client which was to conduct an impact evaluation. Although it may be argued that many indicators for P1 were considered in the use response, it did not clearly present information that could be interpreted as responding to those indicators. The methods and the valuing orientations presented information that more clearly addressed the indicators for the P1 Service Orientation Standard. According to the information provided in each response, it can be concluded that differences in the response provided by the use orientation have been identified via this standard of practice.

2.2.2 Accuracy Standards

The accuracy standards are intended to ensure that an evaluation will reveal and convey technically adequate information about the features that determine worth or merit of the program being evaluated.

A5 Valid Information-- The information gathering procedures should be chosen or developed and then implemented so that they will assure that the interpretation arrived at is valid for the intended use.

Results obtained for the second crucial standard (A5 Valid Information) identified by Stufflebeam (1999), show that the model proposed by the methods orientation was very good (7/10), the model proposed by the valuing orientation was good (6/10) and the model proposed by the use orientation was fair (3/10).

Similarities

All three orientations failed to provide information regarding the two following indicators: Document how information from each procedure was scored, analyzed and interpreted and; report and justify inferences singly and in combination. The valuing and use orientations also failed to provide clear information regarding the documentation and reporting of the data collection conditions and process. The methods and valuing orientations failed to provide clear information regarding training and calibrating the data collectors. The methods and valuing orientations provided similar information in that 9 of the 10 indicators were found to be consistent in both these orientations.

Differences

The information provided by the use orientation also failed to present clear information regarding 8 out of the 10 indicators for this standard (see Appendix G). Since the use response focused on describing an evaluation model (Evaluation Capacity Building) and its implementation, it is not surprising that it would fail to provide information that focuses on *Valid Information*. That is to say, the A5 Valid Information Standard focuses on details regarding documenting, reporting and justifying whereas the use response focused on details regarding ways in which to successfully foster an evaluation culture within a given environment.

Conclusion

Based on the results obtained for the A5 Valid Information Standard, it can be concluded that the methods and valuing orientations obtained similar results. However, differences were observed in the use orientation's response.

A10 Justified Conclusions-- The conclusions reached in an evaluation should be explicitly justified, so that stakeholders can assess them.

Results obtained for the third crucial standard (A10 Justified Conclusions) identified by Stufflebeam (1999), show that the model proposed by the methods orientation was good (6/10), the model proposed by the valuing orientation was fair (4/10) and the model proposed by the use orientation was poor (1/10).

Similarities

All three orientations provided clear information regarding obtaining and addressing the results and prerelease review of the draft evaluation report. However, all three failed to provide clear information regarding the four following indicators: accurately reflect the evaluation procedures and findings; identify and report the program's side effects; explain why rival explanations were rejected and; report the evaluation's limitations. The valuing and use orientations also failed to provide clear information regarding the citing of information that supports each conclusion and the reporting of plausible alternative explanations of the findings. The methods and valuing orientations provided similar information in that 8 of the 10 indicators were found to be consistent in both these orientations.

Differences

The information provided by the use orientation regarding this standard differs greatly from the methods and the valuing response. From the onset of the use orientation's proposal, it was clear that the purpose of the evaluation was to create

and foster an evaluation culture. To do so, the focus of the response was much more on building evaluation capacity to promote ongoing evaluation rather than on the evaluation report. As such, it is not surprising that the use orientation scored poorly regarding this standard. The *Justified Conclusions* (A10) standard is concerned with the presentation of key elements in the evaluation report so as to ensure that conclusions reached are justifiable. Very little information regarding conclusions that would be presented in an evaluation report was provided in the use orientation's response.

Conclusion

According to Stufflebeam's (1999) criteria and indicators, the use orientation failed to provide sufficient information regarding the A10 *Justified Conclusions* standard. Since this standard was identified as crucial, it is recommended by Stufflebeam that the evaluation be failed. However, we are working in the hypothetical and, as such, we can only conclude that differences are identified via this standard of practice. Also, due to the nature of the A10 standard (Justified Conclusions), it was difficult to find clear information in all three responses because the respondents were asked to consider aspects of the planning phase of the evaluation process and the Justified Conclusions Standard is geared toward the outcome of the evaluation. Had they been asked to consider how conclusions could be presented and justified within that which they proposed, results would certainly have been different. Nonetheless, it can still be concluded that differences were observed in the information provided by each response via this third crucial standard.

A11 Impartial Reporting--Reporting procedures should guard against distortion caused by personal feelings and biases of any party to the evaluation, so that evaluation reports fairly reflect the evaluation findings.

Results obtained for the fourth crucial standard (A11 Impartial Reporting) identified by Stufflebeam (1999), show that the model proposed by the methods orientation was good (6/10), the model proposed by the valuing orientation was excellent (9/10) and the model proposed by the use orientation was poor (1/10).

Similarities

As presented in the last table in Appendix G, the methods and valuing orientations obtained similar results on 5 out of the 10 indicators for this standard. The methods and use orientations obtained similar results on 4 of the 10 indicators. The valuing and the use orientations both failed to provide clear information regarding the report of alternative plausible conclusions.

Differences

The most obvious result is that the use orientation failed to provide any clear information regarding the A11 Impartial Reporting Standard. Although some aspects of each indicator could be observed in the use response, each indicator as a whole could not be applied to the information provided in the response. For example, the first indicator which states "Engage the client to determine steps to ensure fair, impartial reports" could be identified in the following information provided by the use response "a mutually agreed-on final report appropriate to the setting" (p.89) however the statement goes on to say "(which is to say, there may not be a lengthy technical document prepared at the study's end)" (p.89). Clearly the intent here is not to safeguard against distortions (as suggested by the indicator) but rather to agree on the length of the report that will be presented.

The indicators for the A11 Impartial Reporting are formulated to safeguard against potential biases. Although the response from the use orientation may touch upon aspects of this, the information always refers back to a different focus than the one intended by the standard.

Another key element that influenced the results of the use orientation regarding this standard is the description of the advisory committee which is made up of carefully chosen staff members who act as "the evaluation "virus" that potentially will "infect" the school's professional community with positive evaluation thinking" (p.90). Furthermore, the intent is to purposefully marginalize certain individuals "Would I include naysayers on this initial committee? […] I advise not including negative people in this initial group. This does not mean, however, that you ignore them; the advisory committee must attend to their interests and concerns individually and extremely purposefully or they may shut the process down" (p.91). Although the "naysayers" may have the opportunity to have their opinion heard, they by no means have any decision-making power in the evaluation process. Only the advisory committee made up of members who are carefully chosen individuals have that power. This could be interpreted as purposefully producing a biased team which may endanger the integrity of the findings.

Conclusion

According to Stufflebeam's (1999) criteria and indicators, the use orientation failed to provide clear information regarding the A11 Impartial Reporting Standard. Since this standard was identified as crucial, it is recommended by Stufflebeam that the evaluation be failed. However, we are working in the hypothetical and, as such, we can only conclude that differences are identified in the three orientations via this standard of practice.

Other data pertaining to the standards of practice also emerged from the findings. Although, not identified by Stufflebeam (1999) as being crucial, these differences are worthy of mention. The following sub-sections will provide insight into these discrepancies.

2.2.3 Feasibility Standards

The feasibility standards are intended to ensure that an evaluation will be realistic, prudent, diplomatic, and frugal.

F1 Practical Procedures--The evaluation procedures should be practical, to keep disruption to a minimum while needed information is obtained.

In their comparative analysis of responses, Alkin and Christie emphasize that King's (use orientation respondent) proposal may not be feasible in the proposed context.

> One of the conditions that King mandates is a substantial (perhaps very substantial) amount of active involvement by school personnel. We wonder about the prevalence of situations where such involvement is possible. Does insistence on active involvement mean that a school like Bunche–Da Vinci would not obtain King's services, and she would do evaluations only in settings where teachers are not so overwhelmed? (p.116)

This reflection was based on Alkin and Christie's intuition since they did not used pre identified criteria in their comparative analyses of the responses. Based upon Stufflebeam's (1999) indicators referring to the F1 standard we support Alkin and Christie's reflection.

To the use orientation's credit however, the information presented in the response does refer to the fact that prior to engaging in the evaluation, a research of the context is necessary to assess whether the environment would be conducive to that which is proposed. However, admittedly "Not every site is interested in such an approach, and given the reportedly negative attitudes of many teachers, this may not be possible at Bunche–Da Vinci" (p.89). Nonetheless, the decision was in favor of signing the contract and proceeding with the evaluation. This decision leads us to question

whether the evaluation proposed by the use respondent is in line with the F1 Practical Procedures Standard.

2.2.4 Propriety Standards

As mentioned (in 2.2.1), the propriety standards are intended to ensure that an evaluation will be conducted legally, ethically, and with due regard for the welfare of those involved in the evaluation, as well as those affected by its results.

P7 Conflict of Interest--Conflict of interest should be dealt with openly and honestly, so that it does not compromise the evaluation processes and results.

As mentioned in a previous section of this chapter, the intent is to train the advisory committee to perform certain activities that other orientations consider to be the role of the evaluator. In this sense, the use respondent proposes to form an internal evaluation team. Many organizations use internal evaluators however, these evaluators are internal to the organization and not to the program being evaluated. Here, the advisory team would be comprised of members who are internal to the program. Perhaps this detail should be shared with the client or it could be considered a violation of this standard.

P8 Fiscal Responsibility--The evaluator's allocation and expenditure of resources should reflect sound accountability procedures and otherwise be prudent and ethically responsible, so that expenditures are accounted for and appropriate.

All three theoretical orientations obtained a low score for the P8 (Fiscal Responsibility) standard. The methods score *Fair* (3/10) and both the valuing and use orientation score *Poor* (2/10 and 1/10 respectively). Clearly none of the three orientations chose to focus on this aspect. Perhaps this is due to priorities within each

orientation or perhaps it is simply a question of space limitations. To assess the true intent on managing the budget and keeping records it would have been necessary to interview the respondents on this issue. Nonetheless, it is interesting that all three orientations provided very little information concerning the Fiscal Responsibility Standard.

2.2.5 Utility Standards

The utility standards are intended to ensure that an evaluation will serve the information needs of intended users.

U6 Report Timeliness and Dissemination--Significant interim findings and evaluation reports should be disseminated to intended users, so that they can be used in a timely fashion.

U7 Evaluation Impact--Evaluations should be planned, conducted, and reported in ways that encourage follow-through by stakeholders, so that the likelihood that the evaluation will be used is increased.

The methods orientation scored *Poor* on both the U6 (Report Timeliness and Dissemination) and the U7 (Evaluation Impact) standards. However, the methods respondent did address this matter in his response.

> Due to space limitations, there are aspects of this case and evaluation plan I was not able to explore or elaborate on in much detail. For example, [...] during step 3, we would have facilitated discussions with the stakeholders to determine how best to disseminate evaluation findings and the lessons learned from the Bunche–Da Vinci evaluation. (p. 82)

Clearly, in a real context, the methods orientation would provide information regarding these two Utility standards. However, in this particular context, respondents were asked to limit the length of their response. As such, choices of what

to focus on were made and, as a result, some details were left by the way side. Nonetheless, mentioning the details that were left out gives us a good sense as to the intent of the response.

To conclude, the data emerging from our sample provides insight into the choice of information each theoretical orientation decided to present in its response. Seemingly, the information provided reflected the priorities inherent to each orientation. In this sense, these results help identify differences existing at the level of the application of the standards of practice as presented by each theoretical orientation.

2.3 Synthesis

Appendix B summarizes the results pertaining to the methodological choices component of the instrumentation which was examined via our first criterion;

1. The presence of *who* would establish each element, *how* this would occur and *when* it will occur.

Findings show little variation regarding *who* is involved in the evaluation process. That is to say, all three orientations agree on involving various stakeholder groups to in the evaluation process. However, a difference was identified at this level. Although, the methods and the valuing orientations selected participants based on similar criteria, the use orientation was found to select participants based on very different criteria.

All three theoretical orientations also provided information regarding *how* data would be collected and interpreted. The methods and valuing orientations used stakeholders in a very similar way. That is to say, stakeholders serve as content or object "experts" in both the methods and valuing orientations. The use orientation was found to view

stakeholder participation in a much broader sense. The primary intended users serve as data collectors and interpreters and, as such, become members of the evaluation staff.

All three orientations set time aside to integrate stakeholder participation into the evaluation process. However, *when* the participation was required varied especially in the use response. The valuing and the methods orientations planned for a few punctual events in which stakeholders would participate. The use orientation required more frequent participation from stakeholders throughout the evaluation process.

Thus, the variations are found at the level of the extent of stakeholder involvement in the evaluation process. In the valuing and methods orientations, the stakeholder involvement is punctual and its purpose is to provide "expert" testimony. The use approach involves a particular stakeholder group (the primary intended users) throughout the evaluation process. Also it requires some stakeholders to collect and interpret data and, as such, relinquishes some aspects considered to be the evaluator's responsibility. In this sense, the main differences among the three theoretical orientations are found through the examination of our second criterion;

2. The evaluator's responsibility in managing *who* would establish each element, *how* this would occur and *when* this would occur.

In applying Stufflebeam's (1999) *Program Evaluation Models Metaevaluation Checklist* (Appendix F) to the three theoretical orientations, results show the responses (models) proposed by the methods and the valuing orientations were quite similar. The following table of results presents the scores obtained for each crucial standard identified by Stufflebeam for each of the three theoretical orientations:

Table 4.9. Scores obtained for each theoretical orientation

Crucial standards identified by Stufflebeam (1999)	Methods orientation	Valuing orientation	Use orientation
P1 Service Orientation	9/10	9/10	5/10
A5 Valid Information	7/10	6/10	3/10
A10 Justified Conclusions	6/10	4/10	1/10
A11 Impartial Reporting	6/10	9/10	1/10
TOTAL SCORE	**28/40**	**28/40**	**10/40**
Average score (sum of scores for each theoretical orientation/n) n=number of crucial standards	**7**	**7**	**2**
Judgment	Very Good	Very Good	Poor

When following Stufflebeam's scale, findings show that on the four crucial standards the methods orientation's proposed evaluation model was very good overall. The valuing orientation's proposed model was also very good overall. The use orientation proposed a model that, according to Stufflebeam, was poor.

Data that emerged serve to confirm specific differences that are seemingly unique to each theoretical orientation except for the P8 Fiscal Responsibility Standard which yielded similar results in all three responses.

2.4 Conclusion of the sub-hypothesis

Results show that our sub-hypothesis, if the modeling is indeed generic and if differences are observed in the evaluation models, then these differences should be at the level of the instrumentation, is confirmed. Confirmation of the sub-hypothesis allows us to confirm our hypothesis in that the discrepancies that emerged are attributed to the instrumentation since the use model deviated at the level of the instrumentation. That is to say, when we applied the instrumentation's methodological component and the standards of practice component to the three theoretical orientations, differences among the theoretical orientations were identified with the most significant difference being identified in the use orientation. As such, we can conclude that the modeling of the practice is indeed generic since differences also occurred at the level of the instrumentation.

CHAPTER V

DISCUSSION

Our research's main focus was to examine the distinction between the specific process of program evaluation and program evaluation models. In doing so, we found that although our hypothesis was confirmed by our global standard it was rejected by our holistic standard.

To investigate whether differences occur at the level of the instrumentation, we conducted a content analysis which focused on the methodological choices and the standards of practice. The most significant difference found in the methodological choices was the intensity of stakeholder involvement required by the model proposed by the respondent representing the use orientation. Furthermore, when we applied the standards of practice to the models proposed by the respondents representing the different theoretical orientations, we found that the methods and valuing orientations yielded similar results in that they both received a rating of *very good*. Significant differences were identified in the use response which obtained a rating of *poor*.

In light of these findings we can safely conclude that our hypothesis was confirmed. That is to say, the holistic standard used to verify our hypothesis showed that the use orientation was the only orientation that did not meet the standard however, when we verified the sub-hypothesis, we found that the use orientation's focus on stakeholder involvement deviated significantly from the process of evaluation to the process of animation. As such, the reason for which the use orientation did not meet the holistic standard is attributed to the instrumentation (the model itself) rather than modeling of the program evaluation process.

Research conducted in the past decades has heavily emphasized evaluation models. Has this emphasis enlightened the practice or has it contributed to the apparent confusion regarding the distinction between theory and practice? Alkin and Christie question whether the use respondent's stance "reflects the autonomy of a university-based evaluator and not that of an evaluator in full-time practice?" (p. 116). These authors also observed that "In discussing the activities proposed by King, we clearly have put the cart before the horse" (p.116). This concern was further developed when Alkin and Christie questioned the applicability of the proposed use oriented model to contexts in which participants are already feeling overwhelmed by the demands of the program.

Since the purpose of their study was not to render a judgment on any individual model or theoretical orientation, Alkin and Christie's observations remained at the level of intuition. However, our frame of reference allows us to look at these concerns from different angles because we gave ourselves a broad frame of reference whose structure includes the modeling of the program evaluation process, methodological choices and standards of practice. This allows us to confirm Alkin and Christie's intuitive concerns with the model proposed by the use respondent. Nonetheless, the use response gives us great insight into the manifestation of the confusion between theory and practice; "To my mind, it is better not to take a contract than to proceed and fail" (King, p. 126). This reflection leads us to question the responsibility of the evaluator. Is it to pick and choose the evaluations in accordance to the model we are comfortable using or to pick and choose the models according to that which is being evaluated?

Results obtained in the present study suggest that the choice of instrumentation should be a function of that which is being evaluated. This brings some clarity to the apparent confusion existing in the literature regarding the relationship between the

evaluation process and the application of the instrumentation. We can recall both Fournier (1995) and Hurteau & Houle (2006) who presented illustrations of the distinction between the program evaluation process and the instrumentation. The illustrations also emphasized the interrelation between these two features by proposing two-way arrows. However, the results obtained in the present research suggest that the representation of this relationship be reconsidered. Since the instrumentation does not influence the modeling of the program evaluation process the relationship between the two may be better illustrated by a one-way arrow stemming from the modeling of the program evaluation process and pointing toward the different components of the instrumentation. This would indicate more accurately the distinction between the two by showing that when the evaluation process is followed the choice of instrumentation will be more appropriate.

Both the methods and the use orientations helped demonstrate this concept. Donaldson (the respondent for the methods orientation) referred to theorists such as Scriven (placed on the valuing branch), Alkin and Patton (placed on the use branch) and Weiss, Rossi and Chen (placed on the methods branch). This mixture of influences and consideration given to all orientations along with the fact that the methods orientation obtained the highest scores demonstrates well the concept of following the modeling of the program evaluation process in order to make appropriate choices regarding the instrumentation. On the other hand, the use orientation insisted on applying an evaluation model that was arguably inappropriate for the evaluation at hand. Since the use orientation's driving factor was the instrumentation (the evaluation model) it deviated from the evaluation process thus demonstrating the necessity of using a generic framework to guide the evaluation process so as not to lose focus on the task at hand (the specific act of evaluating).

At the onset of our research we chose our sample particularly because it created a unique set of data: three different responses to one scenario; ideal in many respects.

However, we do realize that we were dealing in the hypothetical and many factors may have influenced responses causing critical aspects not to be addressed and other aspects to be assumed. Also, we had a very small n. Three responses are hardly enough to establish whether responses were truly representative of their respective theoretical orientation. Perhaps future research could be based on a larger sample of actual evaluation reports conducted by all the theorists presented on Alkin and Christie's (2004) Evaluation Theory Tree. This could allow for patterns to be observed, causal links to be made and conclusion to be drawn regarding the relationship between the modeling of the program evaluation process and the instrumentation.

Future research could also empirically study the extent to which the instrumentation serves the practice. By applying the four components of the instrumentation to evaluations that were conducted and identified as being best practices or failed evaluations, a greater clarity regarding the underlying factors that influence evaluation practices and outcomes could be obtained. These results could contribute to improving the credibility of the field of evaluation by improving the quality of that which it produces.

CONCLUSION

This study set out to contribute empirical data that would be both scientifically and socially relevant. Another objective of the study was to stimulate discussion in order to motivate further research in the practice of program evaluation.

In keeping with this perspective we hypothesized that if the elements of the modeling of the program evaluation process are a generic representation of the program evaluation process, they should be present in the evaluation models. To verify our hypothesis, we applied Hurteau, Lachapelle & Houle's (2005) modeling of the program evaluation process to the three theoretical orientations (methods valuing and use) as identified by Alkin & Christie (2004).

In order to identify whether the discrepancy was at the level of the modeling of the program evaluation process or at the level of the instrumentation, we chose to further orient our hypothesis. As such we formulated the following sub-hypothesis:

> Sub-hypothesis: If the modeling of the program evaluation process
> is indeed generic and if differences are observed in the evaluation
> models, then these differences should be at the level of the
> instrumentation.

To verify our sub-hypothesis we applied the instrumentation (as presented by Hurteau & Houle, 2006) to the three theoretical orientations. The criterion used to guide our investigation of the methodological choices component of the instrumentation was inspired by Alkin and Christie (2005) who, when referring to stakeholder involvement, compared the use response to the valuing response by looking at the *who*, *how* and *when* aspects. As such, we formulated our criteria so as to apply the

methodological choices to each response. We also applied the standards of practice to the three orientations by using Stufflebeam's (1999) *Program Evaluation Models Metaevaluation Checklist* (Appendix F) to the three theoretical orientations. Differences, particularly in the use orientation were identified at the level of the instrumentation. These differences lead us to confirm our hypothesis and allowed us to conclude that there exists a distinction between the process of program evaluation and the instrumentation.

On a social level the study offers insight into the development of program evaluation as a discipline. Results from our study could be useful toward a core body of knowledge. Often evaluation courses present evaluation models as the fundaments of program evaluation. However, the results obtained through the present research show that the fundaments of the practice are the constant elements which define the specific act of evaluating and invariably depict that which is being evaluated. Unlike the different orientations, the fundamental elements are not concerned with *how* an evaluation is conducted it rather focuses on *what* is being evaluated. In order to offer a relevant training course, it would be of great value to teach the invariables that exist in the practice before presenting the different models which vary from one context to the next according to a given theoretical orientation.

Finally, although the present research is limited to three theoretical orientations and is based on a simulated scenario, it has produced valuable empirical data. The data generated allowed us to distinguish between the process of program evaluation and the instrumentation elaborated to support this process. This distinction creates a conscious awareness regarding our responsibility to the field and to the clients we serve. Our primary responsibility is to perform evaluations in all contexts. The instrumentation, including the evaluation models, is a means to this end. Evaluation models as well as the other components of the instrumentation exist as a support to the practice and, as such, they should not dictate our choice of evaluations. The

evaluator's responsibility to the field of evaluation is to make informed decision regarding which instrumentation is appropriate for the context at hand and to apply the fundamental elements of the evaluation process so as to focus on that which is being evaluated in order to generate relevant results and justifiable claims (which has been a weak point of program evaluation practice that regularly resurfaces in debates). We must be very careful not to "place the cart before the horse" or we may become overly concerned with aspects that are external to the program evaluation process and loose sight of our primary responsibility to the field of evaluation and the clients we serve.

REFERENCES

Alkin, M. C. 1990. « Debates on Evaluation ». Newbury Park: Sage Publications.

_____. 2004. « Evaluation Roots: Tracing Theorists' Views and Influences ». Thousand Oaks (Calif.): Sage Publications, 424 p.

Alkin, M.C. and Christie, C. 2005. « Unraveling Theorists' Evaluation Reality ». *New Directions for Evaluation*, no. 106, pp. 111-128.

American Evaluation Association, Task Force on Guiding Principles for Evaluators. (1995). 'Guiding principles for evaluators' in W.R. Shadish, D.L. Newman, M.A.Scheirer, & C. Wye (editors.) « Guiding Principles for Evaluators ». *New Directions for Program Evaluations*. San Francisco: Jossey-Bass, vol. 66, pp. 19-26.

Arens, S.A. 2005. « Understanding Practice through Exemplary Evaluation Cases ». Paper presented at the Canadian Evaluation Society Annual Conference and the American Evaluation Association Annual Conference. Toronto, 2005.

Bouthat, C. 1993. « Guide de présentation des mémoires et thèses ». Montréal : Université du Québec à Montréal, 110 p.

Canadian Evaluation Society. « CES Guidelines for Ethical Conduct ». Available at http://www.evaluationcanada.ca

Chandler, M. 2001. « How Evaluators Engage Theory and Philosophy in their Practice ». Paper presented at the American Evaluation Association Annual Conference. St-Louis.

Chen, H.T. 2005. « Practical Program Evaluation ». Thousand Oaks: Sage publications.

Christie, C.A. 2003. « Understanding Evaluation Theory and its Role in Guiding Practice: Formal Folk and Otherwise ». *New Directions for Evaluation*, vol. 97, pp. 91-93.

Christie, C.A. 2003a. « The Practice-Theory Relationship in Evaluation ». *New Directions for Evaluation*, vol. 97, p. 1-93.

Christie, C.A. 2003b. « What Guides Evaluation? A Study of How Evaluation Practice Maps onto Evaluation Theory ». *New Direction for Evaluation*, vol. 97, pp. 7-35.

Christie, C. and Azzam, T. 2005. « What Theorists Say They Do: A Brief Description of Theorists' Approaches ». *New Directions for Evaluation*, no. 106, pp. 15-26.

Cousins, B.J. and Earl, L. 1992. « The Case for Participatory Evaluation ». *Educational Evaluation and Policy Analysis*, vol. 14, n° 4, pp. 397-418.

Cousins, B.J. and Earl, L. (Eds). 1995. « Participatory Evaluation in Education: Studies in Evaluation Use and Organizational Learning ». London: Falmer.

Cousins, B.J. and Whitmore, E. 1998. « Framing Participatory Evaluation ». *New Directions for Evaluation*, vol. 80, pp.5-23.

Creswell, J. 2004. « Qualitative Inquiry and Research Design Choosing Among Five Traditions ». Thousand Oaks (Calif.): Sage Publications, 403 p.

Datta, L.E. 2003. « Important Questions, Intriguing Method, Incomplete Answers.» In *The Practice-Theory Relationship in Evaluation,* Christina A. Christie (editor). *New Directions for Evaluation,* vol. 97, pp. 37-46.

Donaldson,S.I. 2005. « Using Program Theory-Driven Evaluation Science to Crack the Da-Vinci Code ». *New Directions for Evaluation*, no. 106, pp. 65-84.

Dubois, N. and R. Marceau. 2005. « Un état des lieux théoriques de l'évaluation : une discipline à la remorque d'une révolution qui n'en finit pas ». *The Canadian Journal of Program Evaluation*, vol. 20, n° 1, pp. 1-36.

Duval, J. 2005. « Étude descriptive des pratiques évaluatives : recours à la logique de l'évaluation ». Master of Education Thesis, Université du Québec à Montréal, 79 p.

Eisenberg, N., Winter, L. and Alkin, M.C. 2005. « The Bunche-Da Vinci Learning Partnership Academy ». *New Direction for Evaluation*, no. 106, pp. 5-13.

Eisner, E. 1976. « Educational Connoisseurship and Criticism : Their Form and Function in Educational Evaluation ». *Journal of Aesthetic Evaluation or Education*, vol. 10, p. 135-150.

_____. 1985. « The Art of Educational Evaluation: A Personal View ». Philadelphia (Pen.): The Falmer Press.

_____. 1991a. « The Enlightened Eye ». New York: Macmillan.

_____. 1991b. « Taking a Second Look: Educational Connoisseurship Revisited ». In *Evaluation and Education : At Quarter-Century* (90th yearbook for the Study of Education, Part II), pp. 169-187. Chicago: University of Chicago Press.

_____. 1998. « The Enlightened Eye: Qualitative Inquiry and the Enhancement of

Educational Practice ». Upper Saddle River (NJ): Merrill.

Fetterman, D. 1996. « Empowerment Evaluation: An Introduction to Theory and Practice.» In *Empowerment Evaluation: Knowledge and Tools for Self-Assessment and Accountability,* D.M. Fetterman, S.J. Kafterian, & A. Wandersman (editors). Thousand Oaks (Calif): Sage Publications.

Fournier, D.M. 1995. « Establishing Evaluative Conclusions: A Distinction Between General and Working Logic ». *New Direction for Evaluation,* no 68, p. 15-32.

Guba, E.G. 1972. « The Failure of Educational Evaluation ». In *Evaluating Action Programs: Readings in Social Action and Education,* C. H. Weiss (editor). Boston: Allyn & Bacon. pp. 251-266.

Greene, J.C. 2005. « A Value_Engaged Approach for Evaluating the Bunche-Da Vinci Learning Academy ». *New Directions for Evaluation,* no. 106, pp. 27-45.

Hurteau, M. 1991. « Strategic Choices in Program Evaluation ». In *Evaluation Methods Sourcebook,* A. J. Love (editor). Ottawa (On.): The Canadian Evaluation Society Publications, p.106-125.

Hurteau, M., & Arens, S. 2006. « The impact of credible claims on the practice of evaluation ». American Evaluation Association conference, Portland.

Hurteau, M., Houle, S., & Lachapelle, G. (2006). Comprendre les pratiques en évaluation de programme afin de les améliorer : Un retour aux fondements théoriques. ACFAS, Montréal.

Hurteau, M., Houle, S. 2006. Distinction entre l'acte spécifique d'évaluer et ses modalités de soutien Congrès de l'ADMEE Europe, Luxembourg.

Hurteau, M., Lachapelle, G., & Houle, S. 2006. « La contribution de la logique de l'évaluation dans la pratique évaluation ». *Revue mesure et évaluation en éducation, numéro spécial sur l'évaluation de programme, 29 (3),* 27-44

Hurteau, M., Houle, S., & Duval, J. 2005. « L'étude des pratiques évaluatives : Un retour aux fondements de la discipline ». Joint CES/AEA, Conference, Toronto.

Hurteau M. and Houle S. 2005a. « Identifying a Core Body of Knowledge for Evaluators ». Paper presented at the American Evaluation Association Annual Conference. Toronto, 2005.

Hurteau M. & Houle S. 2005b. « Évaluation de programme et recherche évaluative: des activités distinctes ». *Mesure et Évaluation en Éducation,* vol.28, no. 3, p. 83-95.

Joint Committee on Standards for Educational Evaluation. 1981. « Standards for Evaluations of Educational Programs, Projects, and Materials ». New York: McGraw-Hill.

Joint Committee on Standards for Educational Evaluation. 1994. « The Program Evaluation Standards (2ⁿᵈ ed.) ». Thousand Oaks, (Calif.) : Sage Publications.

Karsenti,T. and Savoie-Zajc, L. 2004. « Introduction à la recherche en éducation ». Sherbrooke : GRP Editions, 350 p.

King, J.A. 2005. « A Proposal to Build Evaluation Capacity at the Bunche-Da Vinci Learning Partnership Academy ». *New Directions for Evaluation*, no. 106, pp. 85-97.

King, J.A., & Stevahn, L. In press. « Interactive Evaluation Practice: Managing the Interpersonal Dynamics of Program Evaluation ». Thousand Oaks, (Calif.): Sage Publications.

Lincoln, Y.S., Guba, E.G. 1985. « Naturalistic Inquiry ». Berverly Hills, (Calif.): Sage Publications.

Madaus, G.F. and D.L. Stufflebeam. 2004. « Program Evaluation: A Historical Overview ». In *Evaluation Models: Viewpoints on Educationnal and Human Services Evaluation* (2ᵉ ed.), D.L. Stufflebeam, G.F. Madaus and Thomas Kellagan (editors), p. 3-19. Boston: Kluwer Academic Publishers.

Mark, MM., Greene, J.C., & Shaw, I.F. 2006. «The evaluation of policies, programs and practices ». In I.F. Shaw., J.C. Greene & M.M Mark, The sage handbook of evaluation, pp.1-30, Thousand Oaks: Sage Publications Ltd.

Mathison, S. and coll. 2004. « *Encyclopedia of Evaluation* ». Thousand Oaks (Calif): Sage Publications, 481 p.

McCarthy, T.A. 1973. « A Theory of Communicative Competence ». *Philosophy of the Social Sciences*, vol. 3, p. 135-156.

_____. 1979. « Communication and Evolution of Society: Jürgen Habermas ». Toronto : Beacon.

McLaughlin, J. and Jordan, G. 1999. « Logic models: A tool for telling your program's performance story ». *Evaluating and Program Planning*, vol. 22, p. 65-72.

Miles, M.B. and Huberman, A.M. 1984. « Qualitative Data Analysis: A Source Book of New Methods ». Beverly Hills, (Calif.) : Sage Publications.

Mongiat, S. 2006. « Études des pratiques en évaluation de programme : retour aux fondements théoriques et à la logique de l'évaluation » Research proposal toward a Master of Education, Université du Québec à Montréal, 59 p.

Owen, J. and Lambert, F. 1998. « Evaluation and the Information Needs of Organizational Leaders ». *American Journal of Evaluation*, vol. 19, pp. 355-365.

Owen, J. and Rogers, P. 1999. « Program Evaluation: Forms and Approaches ». London: Sage.

Owens, T. 1973. « Education Evaluation by Adversary Proceeding ». In *School Evaluation: The Politics and Process*, E.R. House (editor). Berkeley (Calif.): McCutcheon.

Patton, M.Q. 1978. « Utilization-Focused Evaluation ». Beverly Hills (Calif.) : Sage Publications.

_____. 1986. « Utilization-Focused Evaluation (2nd ed.) ». Beverly Hills (Calif.) : Sage Publications.

_____. 1997. « Utilization-Focused Evaluation: The New Century Text (3rd ed.) ». Thousand Oaks (Calif.) : Sage Publications.

_____. 2002. « Qualitative Research and Evaluation Methods (3rd ed.) ». Thousand Oaks, (Calif.): Sage Publications.

Preskill, H. and Torres, R.T. (Eds.). 2001. « The Learning Dimension of Evaluation Use ». *New Directions for Evaluation*, vol. 88.

Redding, P. 1989. « Habermas' Theory of Argumentation ». *Journal of Value Inquiry*, vol. 23, pp. 15-32.

Rossi, P. H., Lipsey, M.W. and Freeman H.E. 2004. « Evaluation: A Systematic Approach ». Thousand Oaks (Calif.): Sage Publications.

Scriven, M. 1967. « The Methodology of Evaluation ». In *Curriculum Evaluation. American Educational Research Association Monograph Series on Evaluation no. 1*, R.E. Stake (editor), Chigago: Rand McNally.

_____. 1972b. « Pros and Cons about Goal-Free Evaluation ». *Evaluation Comment: The Journal of Educational Evaluation*, vol. 3, n° 4, pp. 1-7. UCLA, Center for the Study of Evaluation.

_____. 1974. « Evaluation Perspectives and Proceedures ». In *Evaluation and Education: Current Application*, sous la dir. de J.W. Popham, Berkeley (Calif.) : McCutcheon.

_____. 1980. « The Logic of Evaluation ». Inverness (Calif.) : Edgepress.

_____. 1991. « Evaluation Thesaurus (4th ed.) ». Newbury Park (Calif.): Sage Publications.

_____. 1995. « The Logic of Evaluation and Evaluation Practice ». *New Directions for Evaluation*, n° 68, p. 49-70.

Shadish, W.E., Cook, T.D. and Leviton, L.C. 1991. « Foundations of Program Evaluation Theories of Practice ». Newbury Park: Sage Publications, 529 p.

Suchman, E. 1967. « Evaluative Research : Principles and Practice in Public Service and Social Action Program ». New York : Sage Publications.

Stake, R.E. 2004. « Standards-Based and Responsive Evaluation ». Thousand Oaks (Calif.): Sage Publications.

Stake, R.E., & Schwandt, T.A. (2006). On discerning quality in evaluation. In I.F Shaw., J.C. Greene., & M.M. Mark (Eds.), The *Sage Handbook of Evaluation*, pp. 404-418 Thousand Oaks: Sage Publications

Stufflebeam, D.L. 1983. « The CIPP Model for Program Evaluation ». In *Evaluation Models : Viewpoints on Educational and Human Services Evaluation*, G. F. Madaus, M.S. Scriven et D.L. Stufflebeam (ed.), p. 117-141, Boston : Kluwer-Nijhoff.

_____. 1983. « Lessons in Contracting for Evaluation ». *The American Journal of Evaluation*, vol.21, pp. 293-314.

_____. 1999. « Program Evaluations Metaevaluation Checklist ». *The Checklist Project*. The Evaluation Center. Western Michigan University.

Tellis, W. 1997. « Introduction to case study ». *The Qualitative Report*. On-line series, vol. 3, no.2, 68 pars.

Taylor, P.W. 1961. « Normative Discourse ». Englewood Cliffs (N. J.): Prentice Hall.

Toulmin, S.E. 1964. « The Uses of Argument ». New York : Cambridge University.

Toulmin S.E., R.D. Rieke and A. Janik. 1984. « An Introduction to Reasoning ». New York : Macmillan.

Tyler, R.W. 1942. « General on Evaluation ». *Journal of Educational Research*, vol. 35, p. 492-501.

APPENDIX A

ANALYSIS GRIDS

-Analysis Grid A-

Indicators related to the element *1) Program to be evaluated*

Does the response address the elements of the program under evaluation?	Yes, explicitly	No, it must be deduced or there is no mention of it
1) The clientele is mentioned		
2) The program's objectives are mentioned		
3) The intervention's characteristics are mentioned		
4) The program's theoretical framework is mentioned		

Analysis grids A, B, C, D and E refer to the elements of the modeling of the program evaluation process (Hurteau, Houle, 2006)

-Analysis Grid B-

Indicators related to the element *2) The evaluation's dilemma*

Does the response present the evaluation needs of the client who commissioned the evaluation?	Yes, Explicitly or implicitly (it is possible to deduce since it is evident)	No, it is not specified
1) The evaluation's triggering factor is specified		
2) It is mentioned whether the client expects the evaluator to attribute an order, a score or a mark to the object under evaluation (Type of judgment Scriven,1995)		
3)) It is mentioned whether the client expects the evaluator to support the judgment by basing the evaluation on the judgment of an expert, the causal link between two variables, the perception of the clientele or of the stakeholders, the inherent properties of the intervention, or the needs of the clientele (Criteria sources according to Fournier, 1995)		

-Analysis Grid C-

Indicators related to the element *3) Rendering the evaluation operational*

Does the response attempt to render operational the elements of the dilemma?		
Are the fundamental elements that render the evaluation operational presented?	**Yes, explicitly**	**No, It must be deduced or there is no mention of it**
1) 1) The type(s) of evaluation is/are specified		
2) The question(s) and/or the objective(s) (which specify the type of evaluation according to the evaluation context) is/are mentioned		

Are these elements linked to Rossi, Lipsey, and Freeman's (2004) typology?	**Yes**	**No**	
		The typology is not respected	**Lack of information (see criteria 1 and 2 of this grid)**
3) The type(s) of evaluation and the questions/objectives are linked to Rossi, Lipsey, Freeman's (2004) typology (included in Appendix C)			

Do these elements consider :	**Yes**	**No**	
		They are not considered	**Lack of information**
4) The triggering factor (criterion1 from grid b)?			
5) Criteria presented Scriven or Fournier (criteria 2 and 3 from grid b)			

-Analysis Grid C-

Indicators related to the element *3) Rendering the evaluation operational*

Does the response attempt to render operational the elements of the dilemma?		
Are the fundamental elements that render the evaluation operational presented?	Yes, explicitly	No, It must be deduced or there is no mention of it
1) 1) The type(s) of evaluation is/are specified		
2) The question(s) and/or the objective(s) (which specify the type of evaluation according to the evaluation context) is/are mentioned		
Are these elements linked to Rossi, Lipsey, and Freeman's (2004) typology?	**Yes**	**No**	
		The typology is not respected	Lack of information (see criteria 1 and 2 of this grid)
3) The type(s) of evaluation and the questions/objectives are linked to Rossi, Lipsey, Freeman's (2004) typology (included in Appendix C)			
Do these elements consider :	**Yes**	**No**	
		They are not considered	Lack of information
4) The triggering factor (criterion1 from grid b)?			
5) Criteria presented Scriven or Fournier (criteria 2 and 3 from grid b)			

-Analysis Grid E-

Indicators related to the element *5) Evaluative Claims*

Will a judgment be provided?	
The evaluator-respondent mentions that the intent is not to provide a value judgment on the program's worth or merit. OR The evaluator-respondent provides a description assuming that the program will continue and as such omits to speak about a judgment.	
The evaluator respondent mentions that a value judgment regarding the program's merit will be provided.	
There is no mention of a judgment in the response.	

-Analysis Grid F-

Indicators related to the element *6) Characteristics of the declaration*

How is the argumentation of the evaluative judgment presented? **It is possible to check more than one box**	Yes, prior to rendering a judgment		Yes, at the moment of the judgment		No	
	Implicit	**Explicit**	**Implicit**	**Explicit**	**They are not justified**	**Lack of information (see criteria1 and 2 from grid D)**
1) Does the response justify the choice of criteria? *						
2) Does the response present a justification for the standards?*						
3) Does the response justify other choices?* Specify…						

*In the case where there is a justification, whether it is for the criteria or standards, upon which element of the modeling does the justification relies? *1- Program being evaluated (the client, the program's objectives, the intervention's characteristics, the program's context and its theoretical framework) or 2- The evaluation's dilemma (the evaluation's triggering factor, the type of judgment that the client expects and the argumentation expected by the client)*

Grid F refers implicitly or explicitly to the characteristics of the argumentation

APPENDIX B

SUMMARY TABLE OF COMPARATIVE RESULTS

Summary Table of Comparative Results

Elements of the Modeling	For each orientation, who establishes the elements of the modeling, which methods are used to gather the information and when will this occur?		
	Valuing Orientation	**Methods Orientation**	**Use Orientation**
Program to be evaluated (establishing the existence of the program)	WHO: • evaluation team or consultants • key informants within the school • principal developers of the program HOW: • interviews with stakeholder groups • evaluator reviews available literature WHEN: • at the onset of the evaluation	WHO: • evaluation team made up of experienced members and top level experts HOW: • interviews with various stakeholder groups to elaborate program theories • evaluation team members review plausibility of stakeholder's program theories WHEN: • at the onset of the evaluation	WHO: • Evaluator HOW: • Online research of available information • Questioning acquaintances about what they have heard about the program and the school WHEN: • at the onset of the evaluation
Program's dilemma (identifying the client's preoccupations)	WHO: • various stakeholder groups in collaboration with evaluation team HOW: • interviews with stakeholder groups • evaluation team reviews the needs imparted by way of interviews with various stakeholder groups WHEN: • at the onset of the evaluation	WHO: • various stakeholder groups in collaboration with evaluation team HOW: • interviews with stakeholder groups • evaluation team reviews the needs imparted by way of interviews with various stakeholder groups WHEN: at the onset of the evaluation	WHO: • Principal and evaluator • Participatory evaluation group made up of representatives from different stakeholder groups, a university professor, members of Evaluation Advisory Committee (EAC is a committee if 3 teachers and the principal) and evaluation

			team (2 or more evaluators who h a knowledge of the language, culture, background and required expertise within the context). This evaluation group would meet on a monthly basis. HOW: • First, evaluator and the principal develop a list of proposed project outcomes • Second, serves to identify stakeholder concerns WHEN: • one or more of the monthly meetings with the evaluation group • throughout the evaluations process
Rendering the evaluation operational (formulating evaluation questions and objectives)	WHO: • Evaluation team HOW: • interviews with stakeholder groups • evaluation team reviews the identified priorities imparted by way of interviews with various stakeholder groups • needs are translated into evaluation questions by evaluation team members • the priorities are used as a basis for evaluation team members in the elaboration of evaluation questions and objectives WHEN: • punctual • as needed	WHO: • Evaluation team HOW: • The evaluation team engages relevant stakeholders in discussions about potential evaluation questions • Based on these discussions, evaluation team members formulate evaluation questions WHEN: • punctual • as needed	WHO: • The evaluation group of stakeholder representatives, EAC members and evaluation team HOW: • Different groups engage in discussions and decide upon relevant questions at the time. WHEN: • Regularly scheduled monthly meetings • throughout the evaluations process

Strategic Choices (elaborating the criteria and setting the standards)	WHO: • Stakeholders, in tandem with relevant external perspectives contributed by the evaluator HOW: • established through discussions with diverse stakeholders • needs expressed by the stakeholders are translated into criteria WHEN: • punctual as needed	WHO: • Relevant stakeholders and evaluation team members HOW: • Engage stakeholders to discuss and determine the types of evidence needed to accurately answer the key questions • Stakeholders would rank the evaluation questions in order of priority • Evaluator (and team) would consult various sources of data (i.e. existing performance measures and data sets, document and curriculum review, interview methods, Web-based and traditional survey methods, possibly observational methods, focus groups, and expert analysis). WHEN: • punctual • as needed	WHO: • The evaluation group of stakeholder representatives, EAC members and evaluation team HOW: • Different groups engage in discussions and decide upon relevant criteria at the time. • These groups gather to develop credible methods, help develop instruments, analyze and interpret data WHEN: • Regularly scheduled monthly meetings • throughout the evaluations process
Evaluative claims	WHO: • The evaluator HOW: • The evaluator makes a judgment of the program's quality based on the criteria identified by stakeholders as being of value to the context at hand WHEN: • Punctual at the end of the evaluation process	WHO: • Collaborative process between evaluation team and stakeholders HOW: • reaching agreement on criteria of merit (Scriven, 2003) or agreeing on what would constitute success or failure or a favorable or unfavorable outcome, which will help justify evaluation conclusions and recommendations WHEN: • Punctual at the end of the evaluation process	WHO: • The evaluation group of stakeholder representatives, EAC members and evaluation team HOW: • Group members work collaboratively make recommendations WHEN: • Throughout the evaluation process at regularly scheduled monthly meetings

APPENDIX C

ROSSI, LIPSEY AND FREEMAN'S EVALUATION TYPOLOGY AND MAIN FOCUS
OF EACH EVALUATION TYPE

Rossi, Lipsey and Freeman's Evaluation Typology and Main Focus of Each
Evaluation Type

(Translated from Mongiat, 2006)

Types of evaluations	**Main focus**
Needs assessment	Concerns the targeted clientele's needs and their conditions
Assessment of program theory	Concerns the design and organization of the program
Assessment of program process	Concerns the program's implementation and the services it offers
Impact assessment	Concerns the program's results, effects and impacts
Efficiency assessment	Concerns the program's costs and efficiency

APPENDIX D

SUMMARY OF THE JOINT COMMITTEE'S PROGRAM EVALUATION STANDARDS
(1994)

Summary of the Joint Committee's Program Evaluation Standards (1994)

(prepared by Mary E. Ramlow, the Evaluation Center, Western Michigan University)

Utility Standards

The utility standards are intended to ensure that an evaluation will serve the information needs of intended users.

U1 Stakeholder Identification--Persons involved in or affected by the evaluation should be identified, so that their needs can be addressed.

U2 Evaluator Credibility--The persons conducting the evaluation should be both trustworthy and competent to perform the evaluation, so that the evaluation findings achieve maximum credibility and acceptance.

U3 Information Scope and Selection--Information collected should be broadly selected to address pertinent questions about the program and be responsive to the needs and interests of clients and other specified stakeholders.

U4 Values Identification--The perspectives, procedures, and rationale used to interpret the findings should be carefully described, so that the bases for value judgments are clear.

U5 Report Clarity--Evaluation reports should clearly describe the program being evaluated, including its context, and the purposes, procedures, and findings of the evaluation, so that essential information is provided and easily understood.

U6 Report Timeliness and Dissemination--Significant interim findings and evaluation reports should be disseminated to intended users, so that they can be used in a timely fashion.

U7 Evaluation Impact--Evaluations should be planned, conducted, and reported in ways that encourage follow-through by stakeholders, so that the likelihood that the evaluation will be used is increased.

Feasibility Standards

The feasibility standards are intended to ensure that an evaluation will be realistic, prudent, diplomatic, and frugal.

F1 Practical Procedures--The evaluation procedures should be practical, to keep disruption to a minimum while needed information is obtained.

F2 Political Viability--The evaluation should be planned and conducted with anticipation of the different positions of various interest groups, so that their cooperation may be obtained, and so that

possible attempts by any of these groups to curtail evaluation operations or to bias or misapply the results can be averted or counteracted.

F3 Cost Effectiveness--The evaluation should be efficient and produce information of sufficient value, so that the resources expended can be justified.

Propriety Standards

The propriety standards are intended to ensure that an evaluation will be conducted legally, ethically, and with due regard for the welfare of those involved in the evaluation, as well as those affected by its results.

P1 Service Orientation--Evaluations should be designed to assist organizations to address and effectively serve the needs of the full range of targeted participants.

P2 Formal Agreements--Obligations of the formal parties to an evaluation (what is to be done, how, by whom, when) should be agreed to in writing, so that these parties are obligated to adhere to all conditions of the agreement or formally to renegotiate it.

P3 Rights of Human Subjects--Evaluations should be designed and conducted to respect and protect the rights and welfare of human subjects.

P4 Human Interactions--Evaluators should respect human dignity and worth in their interactions with other persons associated with an evaluation, so that participants are not threatened or harmed.

P5 Complete and Fair Assessment--The evaluation should be complete and fair in its examination and recording of strengths and weaknesses of the program being evaluated, so that strengths can be built upon and problem areas addressed.

P6 Disclosure of Findings--The formal parties to an evaluation should ensure that the full set of evaluation findings along with pertinent limitations are made accessible to the persons affected by the evaluation, and any others with expressed legal rights to receive the results.

P7 Conflict of Interest--Conflict of interest should be dealt with openly and honestly, so that it does not compromise the evaluation processes and results.

P8 Fiscal Responsibility--The evaluator's allocation and expenditure of resources should reflect sound accountability procedures and otherwise be prudent and ethically responsible, so that expenditures are accounted for and appropriate.

Accuracy Standards

The accuracy standards are intended to ensure that an evaluation will reveal and convey technically adequate information about the features that determine worth or merit of the program being evaluated.

A1 Program Documentation--The program being evaluated should be described and documented clearly and accurately, so that the program is clearly identified.

A2 Context Analysis--The context in which the program exists should be examined in enough detail, so that its likely influences on the program can be identified.

A3 Described Purposes and Procedures--The purposes and procedures of the evaluation should be monitored and described in enough detail, so that they can be identified and assessed.

A4 Defensible Information Sources--The sources of information used in a program evaluation should be described in enough detail, so that the adequacy of the information can be assessed.

A5 Valid Information--The information gathering procedures should be chosen or developed and then implemented so that they will assure that the interpretation arrived at is valid for the intended use.

A6 Reliable Information--The information gathering procedures should be chosen or developed and then implemented so that they will assure that the information obtained is sufficiently reliable for the intended use.

A7 Systematic Information--The information collected, processed, and reported in an evaluation should be systematically reviewed and any errors found should be corrected.

A8 Analysis of Quantitative Information--Quantitative information in an evaluation should be appropriately and systematically analyzed so that evaluation questions are effectively answered.

A9 Analysis of Qualitative Information--Qualitative information in an evaluation should be appropriately and systematically analyzed so that evaluation questions are effectively answered.

A10 Justified Conclusions--The conclusions reached in an evaluation should be explicitly justified, so that stakeholders can assess them.

A11 Impartial Reporting--Reporting procedures should guard against distortion caused by personal feelings and biases of any party to the evaluation, so that evaluation reports fairly reflect the evaluation findings.

A12 Metaevaluation--The evaluation itself should be formatively and summatively evaluated against these and other pertinent standards, so that its conduct is appropriately guided and, on completion, stakeholders can closely examine its strengths and weaknesses.

APPENDIX E

CES GUIDELINES FOR ETHICAL CONDUCT

CES GUIDELINES FOR ETHICAL CONDUCT

COMPETENCE

Evaluators are to be competent in their provision of service.

1. Evaluators should apply systematic methods of inquiry appropriate to the evaluation.
2. Evaluators should possess or provide content knowledge appropriate for the evaluation.
3. Evaluators should continuously strive to improve their methodological and practice skills.

INTEGRITY

Evaluators are to act with integrity in their relationships with all stakeholders.

1. Evaluators should accurately represent their level of skills and knowledge.
2. Evaluators should declare any conflict of interest to clients before embarking on an evaluation project and at any point where such conflict occurs. This includes conflict of interest on the part of either evaluator or stakeholder.
3. Evaluators should be sensitive to the cultural and social environment of all stakeholders and conduct themselves in a manner appropriate to this environment.
4. Evaluators should confer with the client on contractual decisions such as: confidentiality; privacy; communication; and, ownership of findings and reports.

ACCOUNTABILITY

Evaluators are to be accountable for their performance and their product.

1. Evaluators should be responsible for the provision of information to clients to facilitate their decision-making concerning the selection of appropriate evaluation strategies and methodologies. Such information should include the limitations of selected methodology.
2. Evaluators should be responsible for the clear, accurate, and fair, written and/or oral presentation of study findings and limitations, and recommendations.
3. Evaluators should be responsible in their fiscal decision-making so that expenditures are accounted for and clients receive good value for their dollars.
4. Evaluators should be responsible for the completion of the evaluation within a reasonable time as agreed to with the clients. Such agreements should acknowledge unprecedented delays resulting from factors beyond the evaluator's control.

APPENDIX F

PROGRAM EVALUATION MODELS METAEVALUATION CHECKLIST

Program Evaluation Models Metaevaluation Checklist

PROGRAM EVALUATION MODELS METAEVALUATION CHECKLIST
(Based on *The Program Evaluation Standards*)
Daniel L. Stufflebeam
1999

This checklist is for performing metaevaluations of program evaluation models. It is organized according to the Joint Committee Program Evaluation Standards. For each of the 30 standards the checklist includes 10 checkpoints drawn from the substance of the standard. It is suggested that each standard be scored on each checkpoint. Then judgments about the adequacy of the subject evaluation model in meeting the standard can be made as follows: 0-2 Poor 3-4 Fair 5-6 Good 7-8 Very Good, 9-10 Excellent. It is recommended that an evaluation model be failed if it scores Poor on standards P1 Service Orientation, A5 Valid Information, A10 Justified Conclusions, or A11 Impartial Reporting. Users of this checklist are advised to consult the full text of The Joint Committee (1994) Program Evaluation Standards Thousand Oaks, CA: Sage Publications.

To meet the requirements for Utility, evaluations using the _____ evaluation model should:

U1 Stakeholder Identification

- Clearly identify the evaluation client
- Engage leadership figures to identify other stakeholders
- Consult potential stakeholders to identify their information needs
- Use stakeholders to identify other stakeholders
- With the client, rank stakeholders for relative importance
- Arrange to involve stakeholders throughout the evaluation
- Keep the evaluation open to serve newly identified stakeholders
- Address stakeholders' evaluation needs
- Serve an appropriate range of individual stakeholders
- Serve an appropriate range of stakeholder organizations

☐ 9-10 Excellent	☐ 7-8 Very Good	☐ 5-6 Good	☐ 3-4 Fair	☐ 0-2 Poor

U2 Evaluator Credibility

- Engage competent evaluators
- Engage evaluators whom the stakeholders trust
- Engage evaluators who can address stakeholders' concerns
- Engage evaluators who are appropriately responsive to issues of gender, socioeconomic status, race, and language and cultural differences
- Assure that the evaluation plan responds to key stakeholders' concerns
- Help stakeholders understand the evaluation plan
- Give stakeholders information on the evaluation plan's technical quality and practicality
- Attend appropriately to stakeholders' criticisms and suggestions
- Stay abreast of social and political forces
- Keep interested parties informed about the evaluation's progress

☐ 9-10 Excellent	☐ 7-8 Very Good	☐ 5-6 Good	☐ 3-4 Fair	☐ 0-2 Poor

U3 Information Scope and Selection

- Understand the client's most important evaluation requirements
- Interview stakeholders to determine their different perspectives
- Assure that evaluator and client negotiate pertinent audiences, questions, and required information
- Assign priority to the most important stakeholders

— Assign priority to the most important questions
— Allow flexibility for adding questions during the evaluation
— Obtain sufficient information to address the stakeholders' most important evaluation questions
— Obtain sufficient information to assess the program's merit
— Obtain sufficient information to assess the program's worth
— Allocate the evaluation effort in accordance with the priorities assigned to the needed information

| ☐ 9-10 Excellent | ☐ 7-8 Very Good | ☐ 5-6 Good | ☐ 3-4 Fair | ☐ 0-2 Poor |

U4 Values Identification

— Consider alternative sources of values for interpreting evaluation findings
— Provide a clear, defensible basis for value judgments
— Determine the appropriate party(is) to make the valuational interpretations
— Identify pertinent societal needs
— Identify pertinent customer needs
— Reference pertinent laws
— Reference, as appropriate, the relevant institutional mission
— Reference the program's goals
— Take into account the stakeholders' values
— As appropriate, present alternative interpretations based on conflicting but credible value bases

| ☐ 9-10 Excellent | ☐ 7-8 Very Good | ☐ 5-6 Good | ☐ 3-4 Fair | ☐ 0-2 Poor |

U5 Report Clarity

— Clearly report the essential information
— Issue brief, simple, and direct reports
— Focus reports on contracted questions
— Describe the program and its context
— Describe the evaluation's purposes, procedures, and findings
— Support conclusions and recommendations
— Avoid reporting technical jargon
— Report in the language(s) of stakeholders
— Provide an executive summary
— Provide a technical report

| ☐ 9-10 Excellent | ☐ 7-8 Very Good | ☐ 5-6 Good | ☐ 3-4 Fair | ☐ 0-2 Poor |

U6 Report Timeliness and Dissemination

— Make timely interim reports to intended users
— Deliver the final report when it is needed
— Have timely exchanges with the program's policy board
— Have timely exchanges with the program's staff
— Have timely exchanges with the program's customers
— Have timely exchanges with the public media
— Have timely exchanges with the full range of right-to-know audiences
— Employ effective media for reaching and informing the different audiences
— Keep the presentations appropriately brief
— Use examples to help audiences relate the findings to practical situations

| ☐ 9-10 Excellent | ☐ 7-8 Very Good | ☐ 5-6 Good | ☐ 3-4 Fair | ☐ 0-2 Poor |

U7 Evaluation Impact
‗ Maintain contact with audience
‗ Involve stakeholders throughout the evaluation
‗ Encourage and support stakeholders' use of the findings
‗ Show stakeholders how they might use the findings in their work
‗ Forecast and address potential uses of findings
‗ Provide interim reports
‗ Make sure that reports are open, frank, and concrete
‗ Supplement written reports with ongoing oral communication
‗ Conduct feedback workshops to go over and apply findings
‗ Make arrangements to provide follow-up assistance in interpreting and applying the findings
☐ 9-10 Excellent   ☐ 7-8 Very Good   ☐ 5-6 Good   ☐ 3-4 Fair   ☐ 0-2 Poor

Scoring the Evaluation for UTILITY Add the following:	Strength of the model's provisions for UTILITY:
Number of Excellent ratings (0-7) _____ x 4 = _____	☐ 26 (93%) to 28: Excellent
Number of Very Good (0-7) _____ x 3 = _____	☐ 19 (66%) to 25: Very Good
Number of Good (0-7) _____ x 2 = _____	☐ 14 (50%) to 18: Good
Number of Fair (0-7) _____ x 1 = _____	☐ 7 (25%) to 13: Fair
Total score: = _____	☐ 0 (0%) to 5: Poor
	_____ (Total score) ÷28 = _____ x 100 = _____

To meet the requirements for feasibility, evaluations using the _____ evaluation model should:

F1 Practical Procedures
‗ Tailor methods and instruments to information requirements
‗ Minimize disruption
‗ Minimize the data burden
‗ Appoint competent staff
‗ Train staff
‗ Choose procedures that the staff are qualified to carry out
‗ Choose procedures in light of known constraints
‗ Make a realistic schedule
‗ Engage locals to help conduct the evaluation
‗ As appropriate, make evaluation procedures a part of routine events
☐ 9-10 Excellent   ☐ 7-8 Very Good   ☐ 5-6 Good   ☐ 3-4 Fair   ☐ 0-2 Poor

F2 Political Viability
‗ Anticipate different positions of different interest groups
‗ Avert or counteract attempts to bias or misapply the findings
‗ Foster cooperation
‗ Involve stakeholders throughout the evaluation
‗ Agree on editorial and dissemination authority
‗ Issue interim reports
‗ Report divergent views
‗ Report to right-to-know audiences
‗ Employ a firm public contract

☐ Terminate any corrupted evaluation				
☐ 9-10 Excellent	☐ 7-8 Very Good	☐ 5-6 Good	☐ 3-4 Fair	☐ 0-2 Poor

F3 Cost Effectiveness

☐ Be efficient
☐ Make use of in-kind services
☐ Produce information worth the investment
☐ Inform decisions
☐ Foster program improvement
☐ Provide accountability information
☐ Generate new insights
☐ Help spread effective practices
☐ Minimize disruptions
☐ Minimize time demands on program personnel

☐ 9-10 Excellent	☐ 7-8 Very Good	☐ 5-6 Good	☐ 3-4 Fair	☐ 0-2 Poor

Scoring the Evaluation for FEASIBILITY Add the following:	Strength of the model's provisions for FEASIBILITY
Number of Excellent ratings 0-3: _____ x 4 = _____	☐ 11 (93%) to 12: Excellent
Number of Very Good (0-3) _____ x 3 = _____	☐ 8 (68%) to 10: Very Good
Number of Good (0-3) _____ x 2 = _____	☐ 6 (50%) to 7 Good
Number of Fair (0-3 _____ x 1 = _____	☐ 3 (25%) to 5 Fair
Total score. = _____	☐ 0 (0%) to 2: Poor
	_____ (Total score) ÷ 12 = _____ x 100 =

To meet the requirements for Propriety, evaluations using the _____ evaluation model should:

P1 Service Orientation

☐ Assess needs of the program's customers
☐ Assess program outcomes against targeted customers' assessed needs
☐ Help assure that the full range of rightful program beneficiaries are served
☐ Promote excellent service
☐ Make the evaluation's service orientation clear to stakeholders
☐ Identify program strengths to build on
☐ Identify program weaknesses to correct
☐ Give interim feedback for program improvement
☐ Expose harmful practices
☐ Inform all right-to-know audiences of the program's positive and negative outcomes

☐ 9-10 Excellent	☐ 7-8 Very Good	☐ 5-6 Good	☐ 3-4 Fair	☐ 0-2 Poor

P2 Formal Agreements, *reach advance written agreements on:*

☐ Evaluation purpose and questions
☐ Audiences
☐ Evaluation reports
☐ Editing
☐ Release of reports
☐ Evaluation procedures and schedule

⎯	Confidentiality/anonymity of data
⎯	Evaluation staff
⎯	Metaevaluation
⎯	Evaluation resources

☐ 9-10 Excellent	☐ 7-8 Very Good	☐ 5-6 Good	☐ 3-4 Fair	☐ 0-2 Poor

P3 Rights of Human Subjects

⎯	Make clear to stakeholders that the evaluation will respect and protect the rights of human subjects
⎯	Clarify intended uses of the evaluation
⎯	Keep stakeholders informed
⎯	Follow due process
⎯	Uphold civil rights
⎯	Understand participant values
⎯	Respect diversity
⎯	Follow protocol
⎯	Honor confidentiality/anonymity agreements
⎯	Do no harm

☐ 9-10 Excellent	☐ 7-8 Very Good	☐ 5-6 Good	☐ 3-4 Fair	☐ 0-2 Poor

P4 Human Interactions

⎯	Consistently relate to all stakeholders in a professional manner
⎯	Maintain effective communication with stakeholders
⎯	Follow the institution's protocol
⎯	Minimize disruption
⎯	Honor participants' privacy rights
⎯	Honor time commitments
⎯	Be alert to and address participants' concerns about the evaluation
⎯	Be sensitive to participants' diversity of values and cultural differences
⎯	Be even-handed in addressing different stakeholders
⎯	Do not ignore or help cover up any participant's incompetence, unethical behavior, fraud, waste, or abuse

☐ 9-10 Excellent	☐ 7-8 Very Good	☐ 5-6 Good	☐ 3-4 Fair	☐ 0-2 Poor

P5 Complete and Fair Assessment

⎯	Assess and report the program's strengths
⎯	Assess and report the program's weaknesses
⎯	Report on intended outcomes
⎯	Report on unintended outcomes
⎯	Give a thorough account of the evaluation's process
⎯	As appropriate, show how the program's strengths could be used to overcome its weaknesses
⎯	Have the draft report reviewed
⎯	Appropriately address criticisms of the draft report
⎯	Acknowledge the final report's limitations
⎯	Estimate and report the effects of the evaluation's limitations on the overall judgment of the program

☐ 9-10 Excellent	☐ 7-8 Very Good	☐ 5-6 Good	☐ 3-4 Fair	☐ 0-2 Poor

P6 Disclosure of Findings

- Define the right-to-know audiences
- Establish a contractual basis for complying with right-to-know requirements
- Inform the audiences of the evaluation's purposes and projected reports
- Report all findings in writing
- Report relevant points of view of both supporters and critics of the program
- Report balanced, informed conclusions and recommendations
- Show the basis for the conclusions and recommendations
- Disclose the evaluation's limitations
- In reporting, adhere strictly to a code of directness, openness, and completeness
- Assure that reports reach their audiences

☐ 9-10 Excellent   ☐ 7-8 Very Good   ☐ 5-6 Good   ☐ 3-4 Fair   ☐ 0-2 Poor

P7 Conflict of Interest

- Identify potential conflicts of interest early in the evaluation
- Provide written, contractual safeguards against identified conflicts of interest
- Engage multiple evaluators
- Maintain evaluation records for independent review
- As appropriate, engage independent parties to assess the evaluation for its susceptibility or corruption by conflicts of interest
- When appropriate, release evaluation procedures, data, and reports for public review
- Contract with the funding authority rather than the funded program
- Have internal evaluators report directly to the chief executive officer
- Report equitably to all right-to-know audiences
- Engage uniquely qualified persons to participate in the evaluation, even if they have a potential conflict of interest; but take steps to counteract the conflict

☐ 9-10 Excellent   ☐ 7-8 Very Good   ☐ 5-6 Good   ☐ 3-4 Fair   ☐ 0-2 Poor

P8 Fiscal Responsibility

- Specify and budget for expense items in advance
- Keep the budget sufficiently flexible to permit appropriate reallocations to strengthen the evaluation
- Obtain appropriate approval for needed budgetary modifications
- Assign responsibility for managing the evaluation finances
- Maintain accurate records of sources of funding and expenditures
- Maintain adequate personnel records concerning job allocations and time spent on the job
- Employ comparison shopping for evaluation materials
- Employ comparison contract bidding
- Be frugal in expending evaluation resources
- As appropriate, include an expenditure summary as part of the public evaluation report

☐ 9-10 Excellent   ☐ 7-8 Very Good   ☐ 5-6 Good   ☐ 3-4 Fair   ☐ 0-2 Poor

Scoring the Evaluation for PROPRIETY Add the following:			Strength of the model's provisions for PROPRIETY	
Number of Excellent ratings (0-8)	_____ x 4 = _____		☐ 30 (93%) to 32:	Excellent
Number of Very Good (0-8)	_____ x 3 = _____		☐ 22 (68%) to 29:	Very Good
Number of Good (0-8)	_____ x 2 = _____		☐ 16 (50%) to 21:	Good
Number of Fair (0-8)	_____ x 1 = _____		☐ 8 (25%) to 15:	Fair
Total score:	= _____		☐ 0 (0%) to 7:	Poor
			_____ (Total score) ÷ 32 = _____ x 100 = _____	

To meet the requirements for accuracy, evaluations using the _____ evaluation model should:

A1 Program Documentation

- ☐ Collect descriptions of the intended program from various written sources
- ☐ Collect descriptions of the intended program from the client and various stakeholders
- ☐ Describe how the program was intended to function
- ☐ Maintain records from various sources of how the program operated
- ☐ As feasible, engage independent observers to describe the program's actual operations
- ☐ Describe how the program actually functioned
- ☐ Analyze discrepancies between the various descriptions of how the program was intended to function
- ☐ Analyze discrepancies between how the program was intended to operate and how it actually operated
- ☐ Ask the client and various stakeholders to assess the accuracy of recorded descriptions of both the intended and the actual program
- ☐ Produce a technical report that documents the program's operations

☐ 9-10 Excellent	☐ 7-8 Very Good	☐ 5-6 Good	☐ 3-4 Fair	☐ 0-2 Poor

A2 Context Analysis

- ☐ Use multiple sources of information to describe the program's context
- ☐ Describe the context's technical, social, political, organizational, and economic features
- ☐ Maintain a log of unusual circumstances
- ☐ Record instances in which individuals or groups intentionally or otherwise interfered with the program
- ☐ Record instances in which individuals or groups intentionally or otherwise gave special assistance to the program
- ☐ Analyze how the program's context is similar to or different from contexts where the program might be adopted
- ☐ Report those contextual influences that appeared to significantly influence the program and that might be of interest to potential adopters
- ☐ Estimate effects of context on program outcomes
- ☐ Identify and describe any critical competitors to this program that functioned at the same time and in the program's environment
- ☐ Describe how people in the program's general area perceived the program's existence, importance, and quality

☐ 9-10 Excellent	☐ 7-8 Very Good	☐ 5-6 Good	☐ 3-4 Fair	☐ 0-2 Poor

A3 Described Purposes and Procedures

- ☐ At the evaluation's outset, record the client's purposes for the evaluation
- ☐ Monitor and describe stakeholders' intended uses of evaluation findings
- ☐ Monitor and describe how the evaluation's purposes stay the same or change over time

⊏ Identify and assess points of agreement and disagreement among stakeholders regarding the evaluation's purposes				
⊏ As appropriate, update evaluation procedures to accommodate changes in the evaluation's purposes				
⊏ Record the actual evaluation procedures as implemented				
⊏ When interpreting findings, take into account the different stakeholders' intended uses of the evaluation				
⊏ When interpreting findings, take into account the extent to which the intended procedures were effectively executed				
⊏ Describe the evaluation's purposes and procedures in the summary and full-length evaluation reports				
⊏ As feasible, engage independent evaluators to monitor and evaluate the evaluation's purposes and procedures				
⊏ 9-10 Excellent	⊏ 7-8 Very Good	⊏ 5-6 Good	⊏ 3-4 Fair	⊏ 0-2 Poor

A4 Defensible Information Sources

⊏ Obtain information from a variety of sources				
⊏ Use pertinent, previously collected information once validated				
⊏ As appropriate, employ a variety of data collection methods				
⊏ Document and report information sources				
⊏ Document, justify, and report the criteria and methods used to select information sources				
⊏ For each source, define the population				
⊏ For each population, as appropriate, define any employed sample				
⊏ Document, justify, and report the means used to obtain information from each source				
⊏ Include data collection instruments in a technical appendix to the evaluation report				
⊏ Document and report any biasing features in the obtained information				
⊏ 9-10 Excellent	⊏ 7-8 Very Good	⊏ 5-6 Good	⊏ 3-4 Fair	⊏ 0-2 Poor

A5 Valid Information

⊏ Focus the evaluation on key questions				
⊏ As appropriate, employ multiple measures to address each question				
⊏ Provide a detailed description of the constructs and behaviors about which information will be acquired				
⊏ Assess and report what type of information each employed procedure acquires				
⊏ Train and calibrate the data collectors				
⊏ Document and report the data collection conditions and process				
⊏ Document how information from each procedure was scored, analyzed, and interpreted				
⊏ Report and justify inferences singly and in combination				
⊏ Assess and report the comprehensiveness of the information provided by the procedures as a set in relation to the information needed to answer the set of evaluation questions				
⊏ Establish meaningful categories of information by identifying regular and recurrent themes in information collected using qualitative assessment procedures				
⊏ 9-10 Excellent	⊏ 7-8 Very Good	⊏ 5-6 Good	⊏ 3-4 Fair	⊏ 0-2 Poor

A6 Reliable Information

⊏ Identify and justify the type(s) and extent of reliability claimed				
⊏ For each employed data collection device, specify the unit of analysis				
⊏ As feasible, choose measuring devices that in the past have shown acceptable levels of reliability for their intended uses				
⊏ In reporting reliability of an instrument, assess and report the factors that influenced the reliability, including the characteristics of the examinees, the data collection conditions, and the evaluator's biases				
⊏ Check and report the consistency of scoring, categorization, and coding				
⊏ Train and calibrate scorers and analysts to produce consistent results				

— Pilot test new instruments in order to identify and control sources of error				
— As appropriate, engage and check the consistency between multiple observers				
— Acknowledge reliability problems in the final report				
— Estimate and report the effects of unreliability in the data on the overall judgment of the program				
☐ 9-10 Excellent	☐ 7-8 Very Good	☐ 5-6 Good	☐ 3-4 Fair	☐ 0-2 Poor

A7 Systematic Information

— Establish protocols for quality control of the evaluation information				
— Train the evaluation staff to adhere to the data protocols				
— Systematically check the accuracy of scoring and coding				
— When feasible, use multiple evaluators and check the consistency of their work				
— Verify data entry				
— Proofread and verify data tables generated from computer output or other means				
— Systematize and control storage of the evaluation information				
— Define who will have access to the evaluation information				
— Strictly control access to the evaluation information according to established protocols				
— Have data providers verify the data they submitted				
☐ 9-10 Excellent	☐ 7-8 Very Good	☐ 5-6 Good	☐ 3-4 Fair	☐ 0-2 Poor

A8 Analysis of Quantitative Information

— Begin by conducting preliminary exploratory analyses to assure the data's correctness and to gain a greater understanding of the data				
— Choose procedures appropriate for the evaluation questions and nature of the data				
— For each procedure specify how its key assumptions are being met				
— Report limitations of each analytic procedure, including failure to meet assumptions				
— Employ multiple analytic procedures to check on consistency and replicability of findings				
— Examine variability as well as central tendencies				
— Identify and examine outliers and verify their correctness				
— Identify and analyze statistical interactions				
— Assess statistical significance and practical significance				
— Use visual displays to clarify the presentation and interpretation of statistical results				
☐ 9-10 Excellent	☐ 7-8 Very Good	☐ 5-6 Good	☐ 3-4 Fair	☐ 0-2 Poor

A9 Analysis of Qualitative Information

— Focus on key questions				
— Define the boundaries of information to be used				
— Obtain information keyed to the important evaluation questions				
— Verify the accuracy of findings by obtaining confirmatory evidence from multiple sources, including stakeholders				
— Choose analytic procedures and methods of summarization that are appropriate to the evaluation questions and employed qualitative information				
— Derive a set of categories that is sufficient to document, illuminate, and respond to the evaluation questions				
— Test the derived categories for reliability and validity				
— Classify the obtained information into the validated analysis categories				
— Derive conclusions and recommendations and demonstrate their meaningfulness				
— Report limitations of the referenced information, analyses, and inferences				
☐ 9-10 Excellent	☐ 7-8 Very Good	☐ 5-6 Good	☐ 3-4 Fair	☐ 0-2 Poor

A10 Justified Conclusions

- ☐ Focus conclusions directly on the evaluation questions
- ☐ Accurately reflect the evaluation procedures and findings
- ☐ Limit conclusions to the applicable time periods, contexts, purposes, and activities
- ☐ Cite the information that supports each conclusion
- ☐ Identify and report the program's side effects
- ☐ Report plausible alternative explanations of the findings
- ☐ Explain why rival explanations were rejected
- ☐ Warn against making common misinterpretations
- ☐ Obtain and address the results of a prerelease review of the draft evaluation report
- ☐ Report the evaluation's limitations

| ☐ 9-10 Excellent | ☐ 7-8 Very Good | ☐ 5-6 Good | ☐ 3-4 Fair | ☐ 0-2 Poor |

A11 Impartial Reporting

- ☐ Engage the client to determine steps to ensure fair, impartial reports
- ☐ Establish appropriate editorial authority
- ☐ Determine right-to-know audiences
- ☐ Establish and follow appropriate plans for releasing findings to all right-to-know audiences
- ☐ Safeguard reports from deliberate or inadvertent distortions
- ☐ Report perspectives of all stakeholder groups
- ☐ Report alternative plausible conclusions
- ☐ Obtain outside audits of reports
- ☐ Describe steps taken to control bias
- ☐ Participate in public presentations of the findings to help guard against and correct distortions by other interested parties

| ☐ 9-10 Excellent | ☐ 7-8 Very Good | ☐ 5-6 Good | ☐ 3-4 Fair | ☐ 0-2 Poor |

A12 Metaevaluation

- ☐ Designate or define the standards to be used in judging the evaluation
- ☐ Assign someone responsibility for documenting and assessing the evaluation process and products
- ☐ Employ both formative and summative metaevaluation
- ☐ Budget appropriately and sufficiently for conducting the metaevaluation
- ☐ Record the full range of information needed to judge the evaluation against the stipulated standards
- ☐ As feasible, contract for an independent metaevaluation
- ☐ Determine and record which audiences will receive the metaevaluation report
- ☐ Evaluate the instrumentation, data collection, data handling, coding, and analysis against the relevant standards
- ☐ Evaluate the evaluation's involvement of and communication of findings to stakeholders against the relevant standards
- ☐ Maintain a record of all metaevaluation steps, information, and analyses

| ☐ 9-10 Excellent | ☐ 7-8 Very Good | ☐ 5-6 Good | ☐ 3-4 Fair | ☐ 0-2 Poor |

Scoring the Evaluation for ACCURACY Add the following:			Strength of the model's provisions for ACCURACY	
Number of Excellent ratings (0-12)	_____	x 4 = _____	☐ 45 (93%) to 48:	Excellent
Number of Very Good (0-12)	_____	x 3 = _____	☐ 33 (68%) to 44:	Very Good
Number of Good (0-12)	_____	x 2 = _____	☐ 24 (50%) to 32:	Good
Number of Fair (0-12)	_____	x 1 = _____	☐ 12 (25%) to 23:	Fair
Total score:		= _____	☐ 0 (0%) to 11:	Poor
			_____ (Total score) ÷ 48 = _____ x 100 = _____	

APPENDIX G

RESULTS OF THE FOUR CRUCIAL STANDARDS INDENTIFIED BY STUFFLEBEAM
(1999)

Results of the Four Crucial Standards Identified by Stufflebeam
(1999)

To meet the requirements for Propriety, evaluation using the ___*___ evaluation model should:			
P1 Service Orientation	**Methods**	**Valuing**	**Use**
Assess needs of the program customers	X	X	X
Assess program outcomes against targeted customers' assessed needs	X	X	X
Help assure that the full range of rightful program beneficiaries are served	X	X	
Promote excellent service	X	X	
Make the evaluation's service orientation clear to stakeholders	X	X	X
Identify program's strengths to build on	X	X	
Identify program's weaknesses to correct	X	X	
Give interim feedback for program improvement	X	X	X
Expose harmful practices	X	X	X
Inform all right-to-know audiences of the program's positive or negative outcomes			

To meet the requirements for Accuracy, evaluation using the ___*___ evaluation model should:			
A5 Valid Information	**Methods**	**Valuing**	**Use**
Focus the evaluation on key questions	X	X	X
As appropriate, employ multiple measures to address each question	X	X	
Provide a detailed description of the constructs and behaviors about which information will be acquired	X	X	
Assess and report what type of information each employed procedure acquires	X	X	
Train and calibrate the data collectors			X
Document and report the data collection conditions and process	X		
Document how information from each procedure was scored, analyzed and interpreted			
Report and justify inferences singly and in combination			
Assess and report the comprehensiveness of the information provided by the procedures as a set in relation to the information needed to answer the set of evaluation questions	X	X	
Establish meaningful categories of information by identifying regular and recurrent themes in information collected using qualitative assessment procedures	X	X	

To meet the requirements for Accuracy, evaluation using the ___*___ evaluation model should:			
A10 Justified Conclusions	**Methods**	**Valuing**	**Use**
Focus conclusions directly on the evaluation on key questions	X	X	
Accurately reflect the evaluation procedures and findings			
Limit conclusions to the applicable time periods, contexts, purposes, and activities	X	X	
Cite information that supports each conclusion	X		
Identify and report the program's side effects			
Report plausible alternative explanations of the findings	X		
Explain why rival explanations were rejected			
Warn against making common misinterpretations	X	X	
Obtain and address the results and prerelease review of the draft evaluation report	X	X	X
Report the evaluation's limitations			

To meet the requirements for Accuracy, evaluation using the ___*___ evaluation model should:			
A11 Impartial Reporting	**Methods**	**Valuing**	**Use**
Engage the client to determine steps to ensure fair, impartial reports	X	X	
Establish appropriate editorial authority		X	
Determine right-to-know audiences		X	X
Establish and follow appropriate plans for releasing findings to all right-to-know audiences		X	
Safeguard reports from deliberate or inadvertent distortions	X	X	
Report perspectives of all stakeholder groups	X	X	
Report alternative plausible conclusions	X		
Obtain outside audit of reports		X	
Describe steps taken to control bias	X	X	
Participate in public presentations of the findings to help guard against and correct distortions by other interested parties	X	X	

*Methods model: Program Theory-Driven Evaluation Science Approach
*Valuing model: Value-Engaged Approach
*Use model: Evaluation Capacity Building Approach