Developmental Evaluation:

The Experience and Reflections of Early Adopters

by

Mark Cabaj

A thesis

presented to the University of Waterloo

in fulfillment of the

thesis requirement for the degree of

Master of Arts

in

Planning

Waterloo, Ontario, Canada, 2011

Author's Declaration

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

The term developmental evaluation first entered in the literature in 1994 when Michael Quinn Patton used the phrase in an article exploring the relationship between evaluation and organizational development. Since then, he has developed the concept further and documented his evolution of thinking in a variety of articles, books and presentations.

While Patton has been systematic in disseminating his emerging account of developmental evaluation, there is very little research on how other evaluators understand the concept nor an account of their experiences using the approach in real-life settings. This thesis documents the experience and reflections of eighteen evaluators who have employed developmental evaluation in their own work. The results suggest that practitioners understand the intent and key features of the approach though have a variety of questions about its conception and encountered a number of practical challenges in its implementation.   The thesis also describes a number of recommendations for strengthening the theory and practice of developmental evaluation.

Acknowledgements

Dedication

This thesis is dedicated to my wife and children: Leann Wagner, Zoe and Isaiah.

Table of Contents

List of Tables

Chapter One: Introduction

1.0 Introduction

The progress of the theory and practice of evaluation has been characterized by an iterative and constant process of reflection and experimentation with ideas and approaches mutating into different forms in diverse circumstances. Alkin (2004) uses the metaphor of an evaluation tree, with roots steeped in the different perspectives of social inquiry and accountability and control, and diverse branches and sub-branches that reflect different approaches to issues such as methods, value and use. House (2004) describes the field as a dynamic river delta with numerous interconnecting pathways.

This thesis research is motivated by a personal desire to contribute to empirical knowledge on evaluation practice as well as an interest in exploring the application of the emerging concept of developmental evaluation. I selected this topic because (a) lack of published research on the topic and (b) my involvement in several learning circles and various interactions with Michael Quinn Patton, the well- known evaluation theorist and practitioner who developed the concept. In this study, I present the results of interviews of "early adopters" of developmental evaluation who shared their experiences and reflections on applying the approach in diverse contexts.

1.1 Overview of Developmental Evaluation

Developmental evaluation entered the evaluation literature in 1994 and 1999 with Michael Quinn Patton's articles on the link between organizational development and evaluation (Patton 1994, 1999). In both documents, Patton describes experiences in which he played an evaluative role in helping organizations create or radically adapt programs and services, rather than the more conventional purpose of improving or judging them. He coined the term developmental evaluation to describe "certain long-term, partnering relationships with clients who are, themselves, in on-going program or organizational development where the role of the

evaluator is to "ask evaluative questions and hold their feet to the fire of reality testing." (Patton 1999, p.109).

In subsequent articles and books, Patton continued to expand and elaborate on the concept, especially the contexts in which evaluative data were useful but required alternative evaluative practices and approaches. These included the development of programs where program designers did not intend to develop a fixed model (Patton 1994, 1999), the process of social innovation and complexity (Westley, Zimmerman and Patton 2006), and replicating and adapting interventions that have proved effective in one context to another (Patton 2008).

Since 1994, Patton has also delivered an estimated 160 workshops, presentations and keynotes on the topic.[1] Evaluators and evaluation users appear to be interested in developmental evaluation. There are on-line communities of practice devoted to the topic (e.g. a Yahoo group), it is discussed regularly on the American Evaluation Society list serve (EvalTalk), over four hundred people signed up for a tele-learning session with Patton on the topic (http://tamarackcommunity.ca/g3s61_VC_2010g.html), and the International Development Research Corporation is developing a curriculum to train evaluators in developmental evaluation so that they can employ it in their work overseas.

In 2008, Patton positioned developmental evaluation as one of six major evaluation purposes, alongside the more traditional evaluation purposes of formative evaluation, summative evaluation, evaluation monitoring, accountability and developing knowledge about patterns of effectiveness (Patton 2008). In 2010, he released a book devoted entirely to the topic of developmental evaluation (Patton 2010).

1.2 The Problem Statement, Research Questions & Rationale

While Patton has been disseminating his thinking on developmental evaluation for some time, and the model appears to be attracting the attention of evaluators and evaluation users, evaluators know very about the underlying concepts and application of the emerging approach beyond the perspective and first-hand experience of Patton. There are no other published

---

[1] Source: Michael Quinn Patton. Conversation on November 5, 2009.

critiques of developmental evaluation, nor articles or case studies reflecting on their use of the approach beyond a "primer" on the topic prepared by a participant-observer community of practice on the topic (Gamble 2008).

For evaluators, this raises the three following questions:

1. What is the current theory of and rationale for developmental evaluation? How does it differ from other forms of evaluation?

2. What are the experiences and reflections of 'early adopters' of developmental evaluation? How do they compare with the emerging theory of developmental evaluation?

3. How can the theory and practice of developmental evaluation be strengthened based on this experience?

The aim of this research is to provide early feedback on the practice of developmental evaluation to evaluation theorists and practitioners in order to inform their understanding and continued development and application of the approach. The purpose of the study is not to judge the fidelity of early adopters' application of developmental evaluation.

1.3 Organization of the Thesis

This thesis is organized around six chapters. In addition to this first chapter, it includes:

Chapter Two: Methodology

This chapter describes the methodology used to investigate the research questions posed in chapter four. This includes key informant interviews with 18 "early adopters" that have employed developmental evaluation in at least one evaluation assignment. It describes the approach to sampling, analyzing and presenting data as well as the limitations of the study.

Chapter Three: Understanding Evaluation

This chapter describes the link between planning and evaluation, and explores a variety of major themes in evaluation theory and practice, including evaluation purposes and issues related methods, design, stakeholders and politics, evaluator capacity and evaluation use.

Chapter Four: Developmental Evaluation

This chapter explores the emerging concept of developmental evaluation based on the existing literature and conversations with Patton, the writing of Jamie Gamble, and related writing on issues related to complexity and innovation. It includes a description of the intent of developmental evaluation and key features of developmental evaluation, the characteristics of developmental situations, related methodologies, the roles and requirements for evaluators, and the limitations and challenges of the concept and its application.

Chapter Five: Results

This chapter describes the patterns and themes emerging from the key informant interview and compares the emerging practice of study participants with the emerging theory of developmental evaluation and surfaces unanticipated questions and themes.

Chapter Six: Conclusions and Recommendations

This chapter describes my conclusions about the research findings, provides a series of recommendations to strengthen the theory and practice of developmental evaluation, as well as a list of new areas for academic research.

Chapter Two: Methodology

2.0 Introduction

This chapter describes the purpose of the study, the major research questions, the research methodology employed, the limitations of the methodology, and the arrangements with the Office of Research Ethics at the University of Waterloo.

2.1 Purpose

The purpose of the study is to determine how early adopters with developmental evaluation understand and apply the emerging theory of developmental evaluation, and how that experience compares with the emerging conceptualization and practice offered by Patton.

The concept and application of developmental evaluation is still emerging, and many of the people have applied it with only limited exposure to Patton's evolving descriptions. Accordingly the results of the study are not being used to judge the fidelity of evaluators' implementation to some as-of-yet-incomplete ideal. It is too early for that type of study.

Instead, the results of the investigation will be used to contribute to the empirical knowledge on evaluation practice and identify implications and recommendations for further refining the emerging theory and practice of developmental evaluation. The study's findings will be of interest to both evaluation theorists and practitioners.

2.2 Research Questions

After presenting the results of a literature review on evaluation and a description of developmental evaluation, this study will explore eight major research questions:

1. To what extent do evaluators distinguish the purpose of developmental evaluation from other evaluations designed to achieve different purposes? How do their experiences compare with the purpose distinctions described by Patton?

2. What criteria and process do evaluators use to determine situations in which DE is appropriate?

3. When do the conditions exist for effective developmental evaluation? How do their experiences compare with the criteria and processes described by Patton?

4. What roles do evaluators fulfill in developmental evaluation assignments? How does this compare to the roles as described by Patton and Gamble?

5. What methods do evaluators use in developmental evaluation? How does this compare with the approach described by Patton?

6. How do evaluators approach accountability in developmental evaluation? How does their experience compare to the dynamic described by Patton?

7. How do evaluators approach the practical task of developing and adapting evaluation designs and budgets? How do their approaches compare to the approach described by Patton?

8. What capacities do evaluators feel they require in order to employ developmental evaluation effectively?

I developed these questions based on two influences. The first are the broader themes related to evaluation theory and practice that I have gleaned from the literature and my own experience. This specifically refers to the major purpose of evaluation research and a number of challenges related to theory and application. The other influence is the questions and observations from other evaluators that I had the opportunity to hear first-hand while attending approximately one-half dozen workshops, presentations and consultations, plus Patton's own thinking on the topic as well as questions and observations from workshop participants.

2.3 Research Paradigms

All research designs are based on an explicit or implicit commitment to at least one of four research paradigms about the nature of reality, experience and knowledge: positivist, naturalistic, critical theory and pragmatic (See Table 2-1). These broad categories in turn hold a number of different orientations and approaches.

This research study is based on a pragmatic paradigm. It employs key informant interviews - a qualitative methodology that is typically used in a naturalistic approach to

Table 2-1: Major Research Paradigms

Paradigm	Epistemology	Purpose	Methods	Criteria
Positivist	Seeks facts or causes of social phenomena apart from the subjective states of individuals.	Generate and test hypothetical-deductive generalizations.	Quantitative and experimental methods.	Internal and external validity, reliability, generalizability.
Naturalistic	Reality and knowledge is inseparable from the persons who experience it.	Inductively and holistically understand human experience and constructed meanings in context-specific settings.	Qualitative and naturalistic.	Confirmability, dependability, credibility and transferability.
Pragmatic	Usable knowledge is situational.	Testing hypothetical-deductive generalizations, inductively understanding phenomenon, abducting	Contingent, menu-based approach to suit context.	Confidence credibility, plausibility and relevance
Critical Theorist	Reality and knowledge is shaped by social structures and language.	Test, challenge and uncover the assumptions underlying perceived experience.	Dialogic methods that combine observation, interviews and reflection.	Based on 'community agreement', acknowledgement of different research paradigms.

Adapted from Robert Woods Foundation Journal (2008) and Patton (2002, page 68-73).

research - not out of an ideological commitment for the paradigm or a personal preference for naturalistic methods, but because it suits the nature of the subject being studied and the purpose of the investigation: i.e. to better understand the experience of evaluators employing an emerging approach to evaluation.

2.4 Interview Research

Researchers use interview research when they want see to "see the world" and gather insight from the perspective of other persons (Ely et al. p. 58). They allow for exploring a smaller "information rich" sample of cases or experiences in more depth (Bert 2007, Zeisel 2006). Kvale (1996, p.2) describes interviews as "a construction site of knowledge" where two or more individuals discuss a "theme of mutual interest" (Kvale and Brinkmann 2009, p. 2). Bogdan and Bilken (2007, p.135) describe it as "… a purposeful conversation usually between people (but sometimes involving more) that is directed by one in order to get information."

2.4.1 Types

There are at least five broad approaches to research interviews: informal conversational interview, the general interview guide approach, standardized open-ended interview, closed and fixed response interview, and the co-constructed or dialogical interview (Patton 2002; Rossman & Rallis 2003).[2] Each strategy has its own inherent strengths and weaknesses.

Informal conversational interviews allow for an emergent conversation between an interviewer and an interviewee. While the conversation is organized around general thematic areas, the interviewer does not determine the questions in advance, but rather allows them to arise naturally during the discussion as salient issues and themes begin to emerge. This provides the interviewee with an opportunity to more clearly express his/her perspectives and experiences on a broad topic, and allows the interviewer to be responsive to the unique context and situation of the interviewee. It is also more difficult to organize and analyze the data arising from informal conversations and the variation in questions, persons and contexts makes it difficult to draw generalizable conclusions. Informal interviews may take place spontaneously as opportunities arise or after considerable advance planning (Marshal & Rossman 2011).

---

[2] There are also at least three specialist approaches: ethnographic interviewing, phenomenological interviewing and focus group interviewing (Rossman et al. 2003).

Table 2-2 Interview Types

TYPE OF INTERVIEW	CHARACTERISTICS	STRENGTHS	WEAKNESSES
Informal Conversational Interview	Questions emerge from the immediate context and are asked in the nature course of things; there is no pre-determination of questions, topics or wording.	Increases the salience and relevance of questions: interviews are built on and emerge from observations; the interview can be matched to individuals and circumstances.	Different information collected from different people with different questions. Less systematic and comprehensive if certain questions do not arise naturally. Data organization and analysis can be quire difficult.
Co-Constructed, Dialogic Interviews	Broad questions determined in advance. Interviewer fully participates in the interview process by sharing their own experiences, reactions and disclosures.	Builds rapport and levels power between interviewer and interviewee. Allows for spontaneous exploration of questions.	Interviewer may lose direction of inquiry. Data organization and analysis can be quire difficult.
Interview Guide Approach	Topics and issues to be covered are specified in advance, in outline form; interviewer decides sequence and wording of questions in the course of the interview.	The outlines increases the comprehensiveness of the data and makes data collection somewhat systematic for each respondents. Logical gaps in data can be anticipated and closed. Interviews remain fairly conversational and situational.	Important and salient topics may be inadvertently omitted. Interviewer flexibility in sequencing and wording questions can result in substantially different responses from different perspectives, thus reducing the comparability of responses.
Standardized Open-Ended Interview	The exact wording and sequence of questions are determined in advance. All interviewees are asked the same basic questions in the same order. Questions are worded in a completely open-ended format.	Respondents answer the same questions, thus increasing comparability of responses; data are complete for each person on the topics addressed in the interview. Reduces interviewer effects and bias when several interviewers are used. Permits evaluation users to see and review the instrumentation used in the evaluation. Facilitates organization and analysis of the data.	Little flexibility in relating to the interview to particular individuals and circumstances; standardized working of questions may constrain naturalness and relevance of questions and answer.
Closed, fixed-responsive Interview	Questions and response categories are determined in advance. Responses are fixed; respondents chooses from among these fixed responses.	Data analysis is simple; responses can be directly compared and easily aggregated; many questions can be asked in a short time.	Respondents must fit their experiences and feelings into the researcher's categories; may be perceived as impersonal, irrelevant, and mechanistic. Can distort what respondents really mean or experienced by so completely limiting their response choices.

Adapted from Daly (2007), Patton (2002), Rossman et al. (2003).

The interview guide approach requires an interviewer to establish the topics and issues of the interview in advance but develop the specific questions, wording and sequence during the interview itself. This allows the interviewer to maintain a conversational tone – which can improve the comfort level of the interviewee – yet allows the interviewer to create an outline of the key topics and themes to be covered in advance, thereby improving the probabilities of a more comprehensive investigation. The variation in interview styles and the emergent approach to questions, however, may result in reduced reliability and comparability of answers by participants (Ely et al. 1991; Marshall & Rossman 2011; Patton 2002).

In a standardized open-ended interview, the questions, wording and sequence of questions are determined in advance and the interviewer asks the same basic open-ended questions in the prescribed order. This allows for a greater comparability of interview responses, permitting interviewees – and evaluation users – to see and review the instrumentation in the research, and thus allows for a more efficient interview process. Its focus on a standardized approach, however, can limit the naturalness of the interview and reduce the flexibility of the interviewer and interviewee to address emergent themes (Ely et al. 1991; Kvale 1996; Marshall et al. 2011; Patton 2002).

In closed, fixed response interviews, the questions and response categories are determined in advance. This is essentially a survey approach that allows interviewers to elicit interviewee responses to fixed responses. It allows for relatively simple gathering and analysis of data but it may be perceived as impersonal by interviewees, distort their experiences and responses by giving them pre-determined options for feedback, and does not allow the interviewer and interviewee to surface and explore emergent themes arising from the interview (Ely et al. 1991; Kvale 1996; Marshall et al. 2011; Patton 2002).

The co-constructed-dialogical interviews recognizes the contributions of both the interviewer and interviewee, aims to have the interviewer and interviewee engage in a process of give, and seeks a nuanced understanding of a topic (Rossman et al. 2003). Interviewers may develop a structured or unstructured set of questions and sequence in advance, through fully

participating in the interview, and by exchanging their own experience, reactions and disclosures (Daly 2007). In this approach, the researcher is an active participant in the co-construction of knowledge, and this encourages interviewees to participate in the active probing and co-generation of knowledge, a process that also "levels" the dynamics of power. The weakness of co-constructed-dialogical interview is that it is often difficult for the interviewer to resist his/her own influences, biases and expectations, which could distort the authentic disclosure of the interviewee's reality (Daly 2007; Rossman et al. 2003)

In practice, these five approaches to interviews could be combined into hybrid strategies. For example, the first part of the interview could employ a closed-fixed response interview strategy while the second part could explore some open ended questions. Moreover, interview strategies are often used as part of a broader strategy of mixed methods, which may include key informant interviews, surveys, direct observation, file reviews, literature review and statistical modelling.

There is no single, best interview strategy. The most appropriate approach is highly dependent on the subject matter, the comfort level of prospective interviewees, the skills of the interviewer, the nature of the relationship between interviewers and interviewees, and the broader research methodology used in the study (Ely et al. 1991; Kvale 1996; Patton 2002).

2.4.2 Design

The design of key informant interviews varies a great deal. Ely et al. (1991) argue that because of the variation in subject matter, researcher preference, and interviewee characteristics, key informant interview design tends to be idiosyncratic, so researchers should be prepared to develop their own defensible designs. Kvale (1996), however, provides a seven step approach that covers most of the issues related to interview research design:

1. Thematizing – formulate the purpose of the investigation and describe the concept of the topic to be investigated before the interviews start.

2. Designing – plan the design of the study, taking into consideration all seven stages, before the interview starts.

3. Interviewing – conduct the interviews with a reflective approach to the knowledge sought.

4. Transcription – prepare the interview material for analysis, which commonly includes a transcription from oral speech to written text.

5. Analyzing – decide, on the basis of the purpose or topic of the investigation, and on the nature of the interview material, which methods of analysis are appropriate.

6. Verifying – ascertain the generalizabiltiy, reliability and validity of the interview findings.

7. Reporting – communicate the findings of the study and the methods applied in a form that lives up to scientific criteria, takes the ethical aspects of the investigation into consideration, and that results in a readable product.

Beyond these general categories of activities, there is a bewildering variety of options that could be employed in each of these steps. Crabtree and Miller (1992), for example, describe four approaches to analyzing qualitative data, ranging from a "pre-figured technical" approach to an "emergent intuitive" approach. These include (a) quasi-statistical analytic style, (b) template analysis, (c) editing analysis style and (d) immersion-crystallization style. Similarly, Guba and Lincoln (1989) identify nine different ways to verify data, including (a) member check, (b) interviewer (b) corroboration, (c) peer debriefing, (d) prolonged engagement, (e) negative case analysis, (f) auditability, (g) confirmability, (h) bracketing, and (i) balance.[3]

Interviewers can complete one round of interviews with one group of interviewees, or, work through several iterations with several groups as part of a continuous, snow-ball investigation where the questions and interviewers from one round informs the next round (Guba et al.1989; Patton 2002).

The general sentiment amongst qualitative researchers appears to be that researchers employing interview methodologies should feel free to develop a contingency approach to design as long as it (a) reflects the broad guidelines of good qualitative research and (b) is carried out ethically (Crabtree et al. 1992; Ely et al. 1996; Marshal et al 2011; Patton 2002).

---

[3] The original source for this reference came from Wikipedia.

2.4.3 Strengths & Weaknesses

In addition to the specific strengths and weaknesses of each interview strategy, there are overarching strengths and weaknesses for the interview methodology itself.

Some of these strengths and weaknesses depend on the research paradigm used by the researcher and/or readers of the research. Positivists, for example, criticize open-ended interviews as "unscientific" because they do not eliminate researcher subjectivity (thereby reducing the validity of the findings) and the responsive method compromises reliability and generalizability. Naturalistic researchers argue that all research paradigms include bias of some sort, and that researchers should simply declare theirs openly and be guided by the equally rigorous standards of confirmability, credibility, transferability, and dependability (Lincoln et al. 1985).

Pragmatists tend to focus on the practical strengths and weaknesses of interviews. They recognize that interviews can be a relatively quick and effective way to gather qualitative data about complex or emergent issues. At the same time, they also point out that interviewees may intentionally and unintentionally provide distorted responses due to personal biases, anger, anxiety, recall error, politics or general lack of reflection or personal awareness (Marshall et al. 2011; Patton 2002). They also highlight the fact that the interaction between the interviewer and interviewee can influence interviewee responses (Ely et al. 1991; Kvale 1996; Patton 2002).

Qualitative researchers have developed an exhaustive list of measures designed to reduce distortions, such as physically locating the interview in a place where the interviewee feels comfortable (Bamberger, Rugh & Mabry 2006) or specialized techniques for interacting with reluctant interviewees (Adler & Adler 2001; Johnson & Wells 2001). Ultimately, however, many conclude that the greatest effect on the quality of the interview is the attributes of the interviewer. This includes his/her ability to frame questions, to listening (or read sign language), inter-personal skills, ability for gentle probing, pattern recognition skills, and verbal communication skills (Ely 1991; Kvale 1996; Marshal et al. 2011).

2.4.4 Ethical Issues

The sensitive nature of working with human subjects, coupled with the desire to gather and analyze accurate and useful data, requires that researchers who use interview techniques must be systematic in their consideration of ethical issues (Chirban 1996; Ely et al. 1991; Gubrium & Holstein 2001; Kvale 1996).

Patton (2002, p. 409) provides an "ethical issues checklist)" that identifies a series of considerations the interviewer should consider throughout the interview process, including: (a) explaining the purpose of the inquiry, (b) promises and reciprocity, (c) risk assessment, (d) confidentiality, (e) informed consent, (f) data access and ownership, (g) interviewer mental health, (h) advice during the researcher, (i) data collection boundaries and (j) other ethical and legal codes used to inform the work.

2.5 Study Research Design

The section reviews sampling strategy, the interview questions and strategy and the approach to analyzing the results.

2.5.1 Questions & Interview Strategy

I originally developed a list of eleven questions based on my reading of the literature and personal experience with developmental evaluation. After I pre-tested the questions with several colleagues with evaluation experience, I reduced the number of questions to eight. I decided to use a structured and open-ended strategy in order to improve the comparability of findings and to improve the probabilities of covering all the study questions.

2.5.2 Sampling & Engagement

While random sampling allows for a greater generalization of research findings and control of selection bias, qualitative sampling allows for a purposeful selection of information-rich cases for more in-depth study.

There are at least fifteen different purposeful sampling strategies that researchers can employ in their research activities: extreme or deviant case (outlier) sampling, intensity sampling, maximum variation sampling, homogenous sampling, typical case sampling, critical case sampling, snowball or chain sampling, criterion sampling, theory based sampling, (dis)confirming cases, stratified purposeful sampling, opportunistic/emergent sampling, purposeful random sampling, sampling politically important cases, convenience sampling, and a combination of mixed purposeful sampling (Patton 2002: 243-244).

This primary sampling strategy used in the study is criterion-based. Criterion based sampling is picking cases that meet some criterion. In this case, the criteria were that interviewees had been (a) exposed to developmental evaluation concepts and practices and (b) that they intentionally employed a developmental evaluation approach in at least one evaluation assignment. Prospective interviewees were identified and recruited from the following sources:

- participants of a Du-Pont-J.W. McConnell Family Foundation sponsored training program on DE with Patton (2005-2006);

- members of the world-wide-web based Yahoo Group devoted to development evaluation;

- a review of articles prepared by evaluators in the *Canadian Journal of Program Evaluation* and the *American Journal of Program Evaluation*;

- participants who reported using developmental evaluation at two workshops on Developmental Evaluation by Michael Quinn Patton Canadian Evaluation Society Annual Conference in 2008 in Ottawa, and participants in another workshop by Cheryl Poth who reported using developmental evaluation in her work.

To enlarge the pool of prospective interviewees, I used snowball/chain sampling to complement criterion-based sampling. This involves identifying prospective study participants from people are aware of prospective interviewees that meet the criteria for the study. In addition to asking prospective interviewees to identify such candidates, I asked Michael Quinn Patton and Jamie Gamble, frequent presenters on developmental evaluation, and Dana Vocisano, the program lead for J.W. McConnell Family Foundation, a large Foundation with many networks in the field of community development and social innovation and the publisher of the Developmental Evaluation Primer, to suggest candidates.

I first approached prospective participants to confirm that they had at least one experience with employing developmental evaluation in their work and to determine their interest in the possibility of participating in the study. I then sent participants who fit these criteria a letter of interest and research waiver form (Attachment 1). Finally, I followed up with people who signed the waiver and established a time for a telephone interview.

I approached twenty-five prospective study participants. Of these, eighteen agreed to participate in the study and seven declined. Of the eighteen people who participated in the study, two were unable to complete the full interview due to time and scheduling constraints.

A summary profile of the interviewees is provided in Table 2-3, which indicates that the profile of study participants is mixed. It includes people who (a) pursue evaluation as a primary focus in their work, a secondary and episodic focus of their work, (b) have formal and/or informal evaluation education and training, (c) years of experience, (d) little to extensive experience with developmental evaluation, and/or (e) different relationships to evaluation users in those assignments.

In addition to the characteristics that I inquired about at the beginning of the interview, the survey process also allowed me to uncover additional characteristics about the interviewees. These include:

- two-thirds of the interviewees are located across Canada. Four were located in the United States, one in New Zealand and one in the Netherlands;

- some respondents were self-professed professional evaluators and non-evaluators using evaluative practices in their work;

- some respondents were managers and administrators responsible for evaluation in their own organization as well as independent consultants who provide evaluation support to a variety of organizations;

- interviewees had experience in fields of education, wildlife management, poverty reduction, public health, environmental protection, disabilities, and the areas of policy change, human services and leadership development; and

- people and organizations operating in the private, public, non-profit and quasi-governmental sector (school boards, health regions) sector.

The diversity provided varied and rich feedback from interviewees.

Table 2-3 Profile of Study Participants

Participant	Evaluation as Focus of Work	Evaluation Education	Relationship To Evaluation users	Years of Evaluation Experience	Exposure to DE	Number of DE Assignments
A	Primary	On the Job	Both	5	DE Training. Reading	3
B	Primary	On the Job	External	30	Reading. Interaction with Patton. Workshop	50 +
C	Primary	On the Job. Prof Dev.	External	15	Reading. Interaction with Patton	NA
D	Episodic	On the Job. Prof Dev.	External	3	DE Training. Reading	1
E	Primary	Masters	External	9	Reading. Interaction with Patton.	6
F	Episodic	Prof. Dev. On the Job	External	15	Reading.	3
G	Primary	Prof. Dev. On the Job	Both	11-12	Reading. Workshop.	2
H	Episodic	DE Training	External	12	DE Training. Reading	1
I	Primary	M.A. Training	External	5	DE Training. Reading. Presentation. Interaction.	13
J	Primary	MA., Phd, Prof. Dev.	Both	20	Reading. Interaction with Patton. Workshop.	2
K	Primary	M.A	External	20	Reading. Interaction with Patton. Workshop	2
L	Secondary	NA	External	2.5 years	Reading. Interaction with Patton. Workshop	1
M	Primary	MA, Phd	Both	20	Reading. Interaction with Patton. Workshop	5[4]
N	Episodic	Prof. Dev. On the Job	External	2.5	DE Training. Reading	2
O	Secondary	Prof. Dev.	Mostly External	?	DE Training. Reading	1
P	Primary	NA	External	7	Reading. Interaction with Patton. Workshop	1[5]
Q	Primary	NA	Internal	10	Interaction with Patton. Workshop	3
R	Primary	M.A.	External	?	Interaction with Patton. Workshop	2

---

[4] This refers to the five developmental evaluations in which the study participant is currently involved.
[5] The participant reported that he was consciously involved in one DE assignment, but many of his past assignments were also likely developmental evaluation.

2.5.3 The Interview Process

The frequency and length of interviews ranged from one interview for ninety minutes to a series of three interviews that cumulatively totalled four hours. While I worked through the interview questions in sequence, in most cases the conversation was emergent and participants would cover themes related to other questions. I allowed the interviewee to continue their response until they felt it was complete and I felt there was no value in probing their response further, and then moved on to the next question in the sequence.

I prepared written notes of each interview and sent interviewees a copy of the notes via email for their review, feedback and sign off immediately after the interview. Once the hard copy of the waiver was signed, I sent the study participant a letter of appreciation via email (Attachment 2).

2.5.4 Analysis

The research analyzed the data from the interviews using a four-step process designed for qualitative research (Miles & Huberman 1984):

1. Tagging – review the raw data from each interview transcripts and use different color highlighters to tag key phrases, sentences and section in the interview.

2. Open Coding – identifying broad themes in the data and assigning initial codes to those themes.

3. Axial Coding – organizing the major themes into sub-themes, and exploring the relationships between those themes and natural clusters.

4. Selective Coding – seeking out selective cases or anecdotes that illuminate themes and allows for easy and accessible comparisons.

Once the data were coded, many of the results were displayed on a series of cross-case display matrices which describe the various ways in which evaluators approach the different aspects of developmental evaluation, and how they compare them with the 'ideal' approach prescribed by Patton. Where and when appropriate, I presented the data in a variety of formats, including matrices, tables and visual diagrams.

In several instances, I reported numerical tallies for coding of interviews. The tallies were not subjected to tests of significance. This is because (a) the small sample size would require nonparametric univariate tests (rather than multivariate analyses) which could lead to overly conservative results, and (b) significant testing encourages dichotomous thinking (e.g. Does it work, yes or no?).

In addition to providing a descriptive comparison of research findings, the analysis also includes extensive narrative descriptions of the data and illustrative quotes from interviewees.

2.6 Limitations

There is a primary and secondary limitation in this study. The primary limitation lies in the choice to rely on a single type of data source (i.e. interviewee perceptions) and single method to gather and analyze data from that source (i.e. an interview)

An ideal research methodology for this study would be a strategy of triangulation, that is, gathering data from multiple sources and using multiple methods to analyze and compare the data from these sources. In order to fully explore the experience of evaluators using developmental evaluation, for example, this would entail interviews with the 'users' of the development evaluation and a file review and content analysis of related evaluation documents.

There are three characteristics of the current practice of developmental evaluation that made such triangulation infeasible within the constraints of this study:

- The volume of evaluative data associated with any developmental evaluation effort is apt to be large (e.g. any "development" decision or action - of any sort in the work of a group or organization), often informal (e.g. memos, minutes of meeting) and frequently not documented (e.g. discussion highlights in a workshop). Given that the study focused on 18 evaluators – some of whom may be basing their responses on their experience with several very large assignment, it would take a great deal effort to analyze and gather written materials and they may not represent an accurate sample of the development findings.

- The nature of the data is often sensitive (e.g. the rationale behind a decision to charge low income residents fees to offset cuts in government funding for a program) and therefore often confidential.

- The number of 'users' and/or participants in a developmental evaluation process is often quite broad, with multiple levels of participation, and turn over frequently as the intervention itself evolves and changes direction.

Despite the reliance on one type of data source and method, the richness of the interviewee's experience and number and diversity of persons interviewed is sufficient to generate useful, information-rich data for the purpose of this study: that is, to better understand the early experiences and reflections of 'early adopters' of developmental evaluation and surface the implications for the concept and its application in the field by other evaluators.

The secondary limitation of this research is its timing. All the interviewees based their understanding of Patton's version of developmental evaluation based on his earlier writings from 1994 to 2008 and participation in various workshops and presentations. I also based my original line of investigation on these same sources.

Patton's comprehensive book on developmental evaluation was released in June 2010, just as I had completed the majority of interviews. While Patton's latest thinking on developmental evaluation is captured in chapter four, many of the questions and concerns study participants raised during the interviews are answered, in whole or in part, in this new book. As a result, some of the experiences, reflections and feedback of interviews are not as relevant as they would have they not already been addressed in these new descriptions.

2.7 Ethics

The University of Waterloo Office of Research Ethics granted ethics approval for the research on February 22, 2010. Research interviewees signed a consent form prior to being interviewed. Copies of the ethics approval form and consent forms are included in Appendix 1 and 2.

Chapter Three: Understanding Evaluation

3.0 Introduction

This chapter provides the context for the research topic. Section 3.1 provides a broad overview of planning and evaluation. Section 3.2 and 3.3 describe the various purposes for which evaluation activities are employed and a number of key concerns in evaluation theory and practice. Finally, Section 3.4 summarizes the main findings in this chapter.

3.1 The Link Between Planning and Evaluation

3.1.1 Definitions

Planning and evaluation are symbiotic. Planners require feedback on the effects of their interventions, the causality between planning interventions and effects, and the degree to which their interventions were successful (Seasons 2002). Evaluators need interventions (planned or otherwise) to evaluate.

While planning and evaluation activities may be easy to spot, there is little consensus in either discipline about how they are best defined. Campbell and Fainstein (2003) argue that defining planning is problematic because (a) it involves larger questions about the role of the state in planning, (b) the practice of planning is trans-disciplinary, bringing together a variety of different professions and methodologies, and (c) planning theorists focus both on the substance of planning (e.g. education, land use planning, etc.) as well as the process (e.g. defining problems, setting priorities, examining trade-offs, etc.).

As a result, planning definitions range from narrow to broad. The two major urban and regional planning associations in North America operate with fairly broad definitions. The Canadian Institute of Planners, for example, defines it as:

Planning means the scientific, aesthetic, and orderly disposition of land, resources, facilities and services with a view to securing the physical, economic and social efficiency, health and well-being of urban and rural communities.

The American Planning Association employs an equally broad definition:

Planning, also called urban planning or city and regional planning, is a dynamic profession that works to improve the welfare of people and their communities by creating more convenient, equitable, healthful, efficient, and attractive places for present and future generations.

These definitions describe the general domain of planning but do not provide insights into its process. While there are different ways to describe the mechanics and steps of planning, as well as variations in the roles of planners that decision-makers and the general public can play in the process, most descriptions tend to include a continuous and iterative cycle of the following activities: identifying and framing problems; establishing, negotiating and deciding on goals, priorities and options for action; and designing, implementing and if appropriate adapting concrete interventions (Etzioni 1967; Mintzberg 1994; Wildavsky 1973; Friedmann 1987).

Defining evaluation is no less challenging. Weiss (1988) defines evaluation as "…the systematic assessment of the operation and/or outcomes of a program or policy, compared to a set of explicit or implicit standards, as a means of contributing to the improvement" (p.4) of the activity. Fournier (2005, p. 140) builds on this idea and provide a similar (if not more elaborate) definition:

Evaluation is an applied inquiry process for collecting and synthesizing evidence that culminates in conclusions about the state of affairs, value, merit, worth, significance, or quality of a program, product, person, policy, proposal, or plan. Conclusions made in evaluations encompass both an empirical aspect and a normative aspect (judgement about the value of something). It is the value feature that distinguishes evaluation from other types of inquiry, such as basic science research, clinical epidemiology, investigative journalism, or public polling.

 Patton (2008) provides perhaps an even broader definition:

… evaluation is the systematic collection of information about the activities, characteristics and results of [interventions] to make judgments about the [intervention], improve or further develop [intervention] effectiveness, inform

decisions about future programming and/or increase understanding. … The definition emphasizes three things: (1) the systematic collection of information about (2) a potentially broad range of issues on which evaluations might focus (3) for a variety of possible judgments and uses (Patton 2008, p.39).

All definitions of planning and evaluation share a common emphasis on the systematic use of knowledge and data in surfacing, developing and assessing public interventions. At the risk of over simplification, planning involves the application of knowledge in the upstream process of (a) defining problems or challenges to be addressed, (b) developing an understanding of the cause-and-effect patterns underlying that challenge, (c) identifying and deciding between possible interventions for changing those patterns, and (d) designing concrete interventions. Evaluation involves the systematic application of knowledge in the downstream activities of (a) gathering and analyzing data on the implementation and effects of these interventions, (b) interpreting the feedback, including (if possible) determining the attribution of these effects and (c) judging the implementation and value of an intervention.

3.1.2 A Brief History of Planning & Evaluation

(a) Pre-Modern Era

People have shown the ability to plan and evaluate their "interventions" well before planning and evaluation became modern disciplines. The survival of hunter-gatherer communities depended on their ability to anticipate the location of game and wild food stuffs and to adapt their tactics when their environment shifted. The emergence of sophisticated agricultural societies in the fertile crescent of the Middle East was only possible due to a relentless process of trial and error with different ways to produce food and domesticate animals. The ability to think evaluatively, plan ahead, and adapt interventions based on feedback is central to technological progress, the emergence of large urban centers and human development in general (Diamond 1999).

There are many institutional examples of planning and evaluation in ancient civilizations. The Egyptians and Aztecs displayed first-rate project management skills when they built the pyramids along the Nile and in Central American jungles, structures that have lasted to the present day. City Administrators metropolis of Carthage experimented with and refined building regulations to ensure that the builders of the city's uniquely tall buildings, the first skyscrapers in

history, were safe and had plumbing systems that met with official standards so that the population growth could contain the city within its high defensive walls. The Emperors of ancient China developed training programs for their civil servants whose proficiency they periodically tested to ensure the ongoing quality of their public service (Patton 2008).

The conceptual DNA of modern planning and evaluation emerged out of the scientific revolution. When astronomers such as Copernicus and Galileo concluded that the earth revolved around the sun (and not the other way around as was commonly assumed) based on a systematic accumulation of data on the movement of the heavens and evolution of alternative theories of the operations of the solar system, they provided one of the most iconic examples of a new epistemology and scientific method at work. At the heart of the paradigm shift, was the simple premise that knowledge about how the world worked needed to be developed using a rigorous process of accumulating evidence and consideration of alternative theories. The implication was that knowledge about reality how the world worked (and any ideas on how it might be changed) was merely a hypothesis that needed to be tested through a battery of experiments.

The scientific revolution helped fuel the renaissance in the 16th century, enabled the Age of Reason in the 17th century and was the engine of the enlightenment in the 18th century. While natural philosophers, such as Frances Bacon and Rene Descartes, elaborated on the methods of the scientific approach, two broad uses of science emerged. The first was "pure science" where scientists employed the experimental approach to uncover the "first principles" of the natural world in such areas as the human body, astronomy, physics, chemistry and biology. They relentlessly documented their findings, and shared and debated them in scientific journals, creating the bedrock of the natural sciences today.

The second manifestation was more utilitarian in nature and laid the seeds for what later would be called applied science and social science: public officials, merchants and craftsmen, were increasingly applying the principles of scientific inquiry to address practical problems in day to day human affairs. City administrators in central London, for example, experimented with new ways to stem the growing number of traffic accidents on the street, eventually settling on a simple but effective regulation which required merchants to drive their lighter empty carts as slowly as when they were loaded with goods for market (Vanderbilt 2009). Close by on the

docks of the Thames, Trading Captain Robert Lancaster became famous for experimenting with different ways to reduce the high mortality rates of the crews on his India-bound trading ships (a drag on profits), including employing a quasi-experimental design in which he gave sailors three daily tablespoons of lemon and compared their scurvy rates with those of their colleagues in the flotilla's other three ships over time (Krippendorf 2008). While these organizational and public policy "experiments" were not carried out as robustly as their natural scientist counterparts, they represented pre-modern manifestations of scientific management and administration.

(b) Early Modern

The industrial revolution created the conditions for the next evolution in planning and evaluation theory and practice. Over a hundred years of progress in natural science enabled the development of technology that harnessed steam power to replace human labour that, when coupled with the factory system which replaced small scale craft production, revolutionized sectors such as agriculture, textiles, manufacturing and transportation. The net result was a dramatic increase material production and economic wealth. It also led to substantial social change and urbanization. Millions of dislocated rural dwellers moved to cities where they found poverty, exploitation of workers, including child labour, which resulted in labour unrest; industrial pollution, overcrowding and poor sanitation, conditions which led to an increase in communicable diseases and high mortality rates; family breakdown and the weakening of traditional community institutions, and higher rates of crime.

These conditions provided city administrators and public officials an opportunity to expand their use of scientific methods in managing human affairs. One of the earliest and most dramatic examples was in the area of public health. In 1855, London authorities became alarmed at the rapid outbreak cholera, one of the deadliest, most infectious and least understood diseases of the time, in a seventy-block area of London. Dr. John Snow, a local physician, had treated patients of smaller outbreak on the London wharves five years earlier and discovered that the majority of victims had shared a common water source. Armed with the hypothesis that victims of the new outbreak were drawing diseased water from one of the community's dozens of public pumps (rather than the more popular theory that they were the victims of 'bad air') he set out to visit as many of the victims and family members as he could to determine where they drew their

water. Snow then recorded the results of each interview on a ledger and created a map with lines connecting the victims to their last known water source. The overwhelming majority of the lines pointed to the main water pump on Broad Street. Twenty-four hours after he convinced local authorities to shut down the pump, only fourteen new cases were reported. Two days later, the death toll stopped entirely. Modern epidemiology was born (Kluger 2008).

Examples of a more scientific approach to human affairs emerged in many areas. In the 1830s, statisticians analyzed a variety of different data sets to test the extent to which education reduced crime in industrial cities in England (Carden 2004). In 1844, civil engineers from Paris used cost-benefit analysis to estimate the value of a canal project by exploring the rates prospective users would be willing to pay (Carden 2004). In the late 1880s and early 1990s, Frederick Winslow Taylor used a stop watch to carry out time-and-motion studies to analyze the "work flow" of different industrial processes in an effort to dramatically improve the productivity of industrial workers. In 1897, Joseph Rue carried out an evaluation of the spelling achievement of over 33,000 primary school students and used the results to help make adjustments to the theory of teaching and the design of school curriculum (Patton 2008).

By the early 20th century, the scientific method to manage human affairs was becoming legitimized. In 1911, Taylor wrote The Principle of Scientific Management, in which he criticized "rule of thumb" management practices and extolled instead the application of principles in studying, designing and managing industrial processes. The same year, the opening address of the first National Conference of City Planners in the United States argued for scientific and orderly planning of cities (Peterson 2003). In 1919, the then future President Wood Wilson laid the groundwork 20th century public administration with his work, The Study of Administration, in which he argued for a more systematic, evidence-based, approach to improving the effectiveness and efficiency of public policy and services.

The field of urban and regional planning was one of the first areas of public administration to embrace the new approach. While many early planners were pre-occupied implementing utopian visions of new cities (e.g. Garden City, City Beautiful), major cities in North America Europe hired reformist planners who applied a more scientific approach to improve existing cities. This included special purpose planning that addressed discrete issues,

such as park design, sanitation or roads as well as comprehensive community plans that created integrated plans for entire cities. From 1900 to 1910, for example, planners in over 80 cities in the United States produced comprehensive city plans, in varying degrees of detail (Peterson 2003). Government legislation and programs to deal with the depression of the 1930s, such as New Deal programs in the United States and Canada, created opportunities for more rigorous public interventions at the Provincial, State and Federal level.

The period between 1900 and 1939 also witnessed the first examples of systematic of evaluation of public interventions. Select foundations and public agencies employed academics to investigate the workings and outcomes of a variety of programs, services and initiatives: e.g. education curriculum and spelling competencies (Patton 2008), public health initiatives to reduce morbidity from infection diseases (Rossi, Freeman & Lipsey 1999), programs to reduce juvenile delinquency and gang activities (Halpern 1994), the effectiveness of settlement houses (Carson 1990), and the implementation and effects of agricultural extension programs (Kirkendall 1966). Several programs, such as the Chicago Area project on juvenile delinquency, were designed and evaluated in the hopes that they might be replicated on a larger scale (O'Conner 1995). These assessments provided opportunities to test and develop inquiry methods better suited to the social – rather than the natural sciences – such as social learning (Dewey 1922), action-research (Lewin 1946), and laboratory research, interviews and questionnaires (Stouffer, Suchman, DeVinney, Williams, Jr. 1949).

Yet, despite the growing and varied incidence of evaluation activities, formal and systematics evaluations were the exception rather than the rule. The mechanisms for public interventions were relatively weak, the theory of evaluation underdeveloped and the number of capable evaluation researchers small. Many of the evaluations that did take place are relatively few, more descriptive in nature and with little information on outcomes (O'Conner 1995; Patton 2008).

(c) High Modernism

The Second World War allowed central governments to accumulate an unprecedented degree of authority, resources and skills in administration and management. Once the war was over, the general public and politicians turned their attention to building "Great Societies" back home. The dramatic expansion of the welfare state in most industrialized and industrializing countries involved a large number and variety of interventions into areas such as urban renewal, education, health and economic development. This in turn created an enormous demand (and eventually capacity) for a systematic approach to planning and evaluating public interventions based on rigorous social science.

The planning manifestation of high modernism was rational comprehensive planning. It represented an extreme (albeit logical extension) of the scientific method. Faced with a problem, such as whether to tear down a distressed neighborhood and start anew, planners would ideally list all the relevant variables related to that issue (e.g. new tax revenues, the cost of borrowing, the expropriation of land, effect on traffic), rate these variables in terms of their importance, possibly even accommodating the different value placed on each variable placed by diverse stakeholders, and then develop a range of possible alternative interventions and their possible outcomes, using whatever theories available to him/her by social science. If possible, the results of each alternative and outcome scenario would include a cost-benefit analysis, which provided a detailed monetary account of the benefits and costs with each alternative. Based on a rigorous assessment of all alternatives, the planner would select the alternative that reflected the policy priorities, offering the result of the deliberation to public decision-makers.

The role of evaluation was to provide feedback on – and judge the worth of – these interventions using equally rigorous scientific methods. This involved the extensive use of quantitative methods (e.g. randomized controlled trials) to measure the effects and causality of interventions and elaborate techniques for judging the worth of interventions (e.g. rating scales, cost-benefit analysis). Unlike the previous generation of evaluation, evaluators periodically made the extra effort to go beyond whether programs achieved their goals, but also to search for unintended effects, or externalities, of an intervention.

Public administrators took the planning and evaluation functions so seriously that they institutionalized both in legislation and public agencies. Following the lead of the United States, the Government of Canada instituted a Planning, Programming and Budgeting System in the 1960s. In 1973, it established an evaluation branch in Office of the Comptroller General in 1973 and in 1976 mandated all government departments to include performance measurement plans and periodic reviews of their programs.[6] The greater demand for planning and evaluation in turn helped fuel a significant expansion in the professional infrastructure in both fields: universities developed courses in both disciplines, think tanks and policy institutes created evaluation departments, and professionals established associations, journals and publications.

Despite its hopeful start, the high modernist period in planning and evaluation was relatively short-lived as practitioners, theorists and administrators uncovered a long list of limitations of both rational comprehensive planning and an evaluation approach rooted in social science. This included the following critiques: (a) that the technical, financial and intellectual requirements of outstripped the capacity of most public institutions (Wildavsky 1973) and human intellect in general (Simon 1957), (b) that the assumption of a common public good was false and in reality the public good was made up of a pluralist public operating with varied – often conflicting - values and interests of a diverse society (Fainstein 2003), and (c) planners consistently ignored feedback on the counterproductive effects of their intervention because it did support their own visions of how cities societies did (and should) work (Jacobs 1961).

At a more fundamental level, critics of RCP and RCTs argued that the very epistemological paradigm underlying the approaches – the positivist paradigm inherited from the natural sciences – was unsuitable for the complex realities of the social world. Constructivists (also known as interpretivists and naturalists), argued that reality in the physical world was objective, in the social world it was "subject dependant". The role of planners and evaluators, therefore, was to help surface these different experiences through communicative planning and naturalistic evaluation (Fainstein 2003; Guba et al. 1989). Critical theorists went a step further when they claimed that peoples' understanding of the world was so deeply shaped by language, culture and systems that planners and evaluators needed to actively involved in uncovering (and keeping in mind) those biases, such as gender (Vainio-Mattilo 1999) and race (The Aspen

---

[6] See Ian Greene at http://www.yorku.ca/igreene/progeval.html

Institute 2005). Pragmatists found themselves somewhere in the middle of the debate, arguing that some parts of reality were objective and some were subject-dependant, and that planners and evaluators had to take both into account (Pawson & Tilley 1997).

In the end, rational comprehensive planning and randomized controlled trials were the exception and not the general pattern. In his article, If Planning is Everything, Maybe It's Nothing, Wildavsky (1973) described finding very few examples of consistently applied comprehensive planning, a research finding confirmed by Dalton (1986) and Mintzberg (1994) decades later. Similarly, very few evaluations managed to meet the exalted gold standards of randomized controlled trials. Stufflebeam (2004) describes how as early as the 1960s, education evaluators concluded that they would be able to employ RCT's in only a few rare cases. Cook (1993), a well-known advocate of RCTs, reported that he found only a few examples of text-book investigations over his time as an evaluator. Shadish (Shadish et al. 1991; 2002), another experienced social scientist, reported finding only six such examples in his career (Shadish, Cook and Leviton 1991; Shadish, Cook and Campbell 2002).

(d) Post-Modernism

The dissatisfaction with high modernism created the space for the post-modernist period which was characterized by wide-spread experimentation and development of other planning and evaluation modalities and epistemological paradigms. Some of the better known planning approaches include: incrementalism (Lindblom 1959), mixed scanning (Etzioni 1967), advocacy planning (Davidoff 1964), participatory planning (Arnstein 1969), contingency planning (Thompson 1967;, Alexander 1984), communicative planning (Fainstein 2003), and feminist planning (Greene & Caracelli 2007). In the area of evaluation, some of the new methods included: adversarial evaluation (Wolf 1975), responsive evaluation (Stake 1974), deliberative democratic evaluation (House & Howe 1999), naturalist evaluation (Guba et al. 1989), participatory evaluation (Cousins 1994), empowerment evaluation (Fetterman 2001), realist evaluation (Pawson & Tilley 1997) and feminist evaluation (Seilbeck-Bowen, Brisolara, Tischler, Whitmore 2002). While the field was developing new approaches, high modernists continued to elaborate on the positivist approaches, including new social science methodologies

such as theories of change (Weiss 1972), quasi-experimental designs (Campbell and Stanley 1973) and social return on investment (Emerson 2000).

The vitality of planning and evaluation in the post-modern period coincided with (and in part caused) confusion and a lack of confidence in both fields. In 1973, Anton Wildavsky, the celebrated policy analyst and keen observer of planning, summarized the general state of planning:

> The planner has become the victim of planning; his own creation has overwhelmed him. Planning has become so large that the planner cannot encompass its dimensions. Planning has become so complex planners cannot keep up with it. Planning protrudes in so many directions, the planner can no longer discern its shape. He may be economist, political scientist, sociologist, architect or scientist. Yet the essence of his calling--- planning-- escapes him. He finds it everywhere in general and nowhere in particular. Why is planning so elusive? (Wildavsky 1973: 127).

He went on to describe the effects of failed planning on the planners themselves:

> Planners can no longer define a role for themselves. From old American cities to British new towns, from the richest countries to the poorest, planners have difficulty in explaining who they are and what they should be expected to do. If they are supposed to doctor sick societies, the patient never seems to get well. Why can't the planners ever seem to do the right thing?

Not everyone agreed. Alexander (1981), for example, found Wildavsky's argument overly polemic and provided an elaborate counter-argument to each of his claims. But he seemed to represent a minority position. In a cynical moment, John Friedmann (1966) argued that many dispirited planners and academics had begun to view planning simply as a way to make a living rather than as vocation in service of some higher aspiration for the public good.

Evaluators appeared to be in a similar dis-array. In 1982, Palumbo and Nachmias described the angst within the profession of evaluation, brought on in part by the rapid expansion of its aims, methodologies and roles:

> The field of evaluation is undergoing an identity crisis. From its initial surge in the 1960s when evaluation research clearly was dominated by a single methodology and evaluation researchers believed that is potential

was unlimited, it has undergone a metamorphosis. Rather than a single orientation, a number of alternative approaches to evaluation have sprung up and a nagging doubt about its future has crept into a number of recent publications. The scale, ubiquity, and diversity of evaluation activities made comprehension difficult, even for those operating within the field".

Ernest House echoed the sentiment, noting that the "current evaluation scene is marked by vitality and disorder"[7].

(e) Planning & Evaluation Today

There appears to be very little substantial development in the theory of planning and evaluation in the last several years. The proliferation of resources and techniques appear to elaborate on existing approaches rather than represent radically new ways. In his introduction to urban planning assessment of modern planning in the introduction to his text book on the topic seems appropriate: "there are just a few key ideas in twentieth century planning which re-echo, recycle and reconnect" (Hall 2002; 7). The pattern in the evaluation field looks roughly similar: none of the twenty-four evaluation theorists in Alkin's (2004) evaluation reader, for example, identified the emergence of radical new thinking in evaluation.

The exception is a simmering continuation of the "paradigm war" that began in the high modernist period (Bamberger, Rught and Mabry 2006). Chambers (2010) summarized the tension as a debate between neo-newtonians, who perceive the world through a positivist lens, and like "regular, linear, predictable" governed by universal laws that can be revealed through methods developed through the natural science. On the other end of the spectrum are "adaptive pluralists", people who view the world as "non-linear, emergent and unpredictable" where social problems are context sensitive, defying standard universal laws, and reality is at least in part "subject-dependant". In some cases, the positions are sufficiently polarized that organizations and practitioners in each field have opted to commit fully to one paradigm over another rather than adopt a centrist, pragmatist position that characterizes most of the field (Chambers 2010; Patton 2008).

---

[7] Ernest House, quoted in Dennis J. Palumbo and David Nachmias, "The Pre-Conditions for Successful Evaluation: Is there an Ideal Type?", paper prepared for presentation at the International Political Science Association Meetings, Rio de Janeiro, Brazil, August 9-14, 1982, p. 1; subsequently published in Policy Sciences 16 (1983): 67-69 and Implementation (1983), Pressman and Wildavsky: page 182.

3.2 Evaluation Purposes

The dynamic evolution of evaluation has resulted into a broad and complex field. There are approximately eighty different evaluation methods and techniques (Evaluation Wiki 2011), a variety of epistemological paradigms about what constitutes knowledge and different disciplines, such as education, anthropology and the natural sciences (Alkin 2004) and countless practical factors that shape the selection and design of methodsvariations in interventions, contexts, and users that shape each evaluation (Bamberger et al. 2006; Stake 1974; Stufflebeam 1968). If the "vitality and disorder" is confusing to evaluators, it must be more so for evaluation users (Pressman et al. 1983).

Ultimately, it is possible to reduce this complexity to a manageable size if evaluations are organized around different purposes. Over the years, roughly six major discrete purposes have emerged. These include: summative evaluation, formative evaluation, pre-formative, accountability, monitoring, and generating knowledge. Table 3-1 describes the major characteristics of each purpose, as well as typical questions, users and methods employed for each of them.

The following section briefly describes the evolution and aim of each purpose in turn.

3.2.1 Summative

The earliest niche for evaluation to emerge was to help public officials and those investing in interventions to judge their merit or worth. The rapid growth in the number and variety of public and private programs and policies in industrialized countries created a demand for information to help decisions makers answer a variety of questions: Do these interventions work? Are they worth it? Should they be continued or replicated? (Scriven 1991, p. 240) provides, the person to popularize the phrase summative evaluation, describes it as follows:

Table 3-1 Evaluation Purposes[8]

PURPOSE	PRIORITY QUESTIONS	TYPICAL USERS	COMMON METHODS
Summative	Does the program meet peoples' needs? How does its cost compare to its benefits? To what extent are the outcomes attributable to program activities? Should we sustain this, drop it, scale it up or replicate it?	Funders; those charged with making major decisions about the program's future (e.g. board of directors, policy makers).	Impact evaluation. Cost-benefit analysis. Theory driven evaluation.
Formative	What works and what doesn't? What are program participants saying? How does this differ for different groups? How can we increase program effects? Decrease costs?	Program administrators, staff, and participants; those involved in the day to day operations.	Quality improvement. Learning reviews. Appreciative inquiry. Reflective practice. Participant feedback.
Pre-Formative	What is the problem we are trying to define? What are possible solutions and likely effects? What types of mechanisms for change are required? How feasible are they?	Social innovators, policy designers who assume lead roles in designing a program.	Delphi Techniques Simulations Modeling
Monitoring	Are inputs and processes flowing smoothly? What are participant dropout rates? Are they changing? Are outputs being produced as anticipated and scheduled? Where are the bottle-necks?	Program managers of day to day operations; higher level decision-makers interested in linking monitoring to planning and budgeting cycles.	Management information systems. Quality control systems. Routine reporting. Performance indicators.
Knowledge Generating	What are the general patterns and principles of effectiveness across program sites? What lessons are being learned? How do evaluation findings triangulate with research results, social science theory, expert opinion, participant feedback? What principles can we extract to inform practice?	Program designers, planners, modellers, theorists, scholars and policy makers.	Cluster evaluation. Meta-analysis. Synthesis evaluation. Lessons learned. Effective practice studies.
Accountability	Are funds being used for intended purposes? Are goals and targets being met? Is staff qualified? Is implementation following the plan? Are only eligible participants being served?	Those with executive managerial, legislative, and funding authority to ensure scare resources are used well.	Mandated reporting. Program audits and inspections. Accreditation and licensing. End of project reports. Scorecards.

---

[8] This table has been adapted from Patton 2008: 139-141.

Summative evaluation of a program (or other evaluand) is conducted after completion of the program (for ongoing programs that means after stabilization) and for the benefit of some external audience or decision-maker (for example, funding agency, oversight office, historian, or future possible users) …. The decision it services are most often decisions between these options: export generalize), increase site support, continue site support, continue with conditions(probationary status), continue with modifications, discontinue …. The aim is to report on it, not to report to it.

In order to make these decisions, policy makers and funders require answers to more detailed evaluation questions: i.e. What are the effects of this intervention? To what extent is the intervention is responsible for these changes and not some other factors? What are the criteria used to judge the merit or worth of this intervention?

Given the high stakes of making a decision about the merit and future of an intervention, debates about epistemological orientation, methodological approach and the role of the evaluator loom large in summative evaluation. Advocates of a positivist approach favour the use of external evaluators and a battery of quantitative techniques (e.g. cost-benefit analysis and statistical modeling), including randomized control trials, to measure the incremental effects of an intervention and to eventually judge it against a set of pre-determined criteria (Campbell & Stanley 1973). Evaluators who prefer naturalistic or constructivist approaches, on the other hand, stress the importance of developing "thick qualitative descriptions" of the observed effects and judgements of a diverse set of intervention stakeholders (Guba et al. 1989). Pragmatists typically argue for a mixed methodology strategy able to provide the kind of evidence a decision-maker(s) requires and is feasible given the context and constraints in which it was being employed (Schorr 1997).

In practice, evaluators and evaluation theorists often find it difficult to generate the information they would like to make high stake summative decisions. There are a variety of reasons for this, including among others: the data on the intervention effects is weak and/or difficult to measure, the attribution of the effects to the intervention uncertain, the criteria used to judge the merit or worth of the intervention contested, and the intervention being judged continues to evolve (Patton 2008). These challenges exist for relatively discrete interventions, such as a labour market training programs (Hum and Simpson 2002), programs to lend small businesses money for job development (Watson 1994) as well as complex and open-ended

interventions, such as comprehensive initiatives that aim to create population-level changes in distressed neighbourhoods (Aspen Institute 1995, 1997a, 1997b).

3.2.2 Formative

The second dominant niche for evaluation is formative evaluation. While summative evaluations aim at judging the merit or worth of an intervention after it's had a sufficient opportunity to yield some effects, formative evaluations are used to improve the design and delivery of an intervention once its implementation begins (Cronbach 1980; Scriven 1967, 1991). The primary users of formative assessments, therefore, tend to be the designers and administrators of interventions, rather than those who make political or resource decisions about the intervention.

Formative evaluation questions tend to be similar regardless of whether the intervention being evaluated is in its early or mature phases of implementation: Are things going to plan? What are the beneficiaries of the intervention saying about the program? What are the strengths and weaknesses of the model? Where and how might we adjust the design or implementation of the intervention in order for better performance?   (Bamberger, et al. 2006; Patton 2008; Stufflebeam & Webster 1981; Stufflebeam 2004).

Over time, the emphasis on improving interventions has spawned many modes of inquiry and management in the private, public and non-profit sector. These include, among others, total quality management, six-sigma, and continuous quality improvement (Patton 2008).

There is a close and symbiotic relationship between formative and summative evaluation. At the height of Great Society programs, Scriven (1967) described how formative evaluation would help program designers and administrators work out the wrinkles of promising policy and program interventions. Once the program had stabilized, summative evaluation would then be employed by policy makers and funders to discover the effects of the program and judge the overall merit of the intervention, deciding whether it should be sustained, discontinued or scaled up.   Some evaluators called the combination of summative and formative evaluation "comprehensive evaluation" (Freeman 1977).

3.2.3 Monitoring

Evaluation monitoring represents a third niche for evaluation. Unlike pre-formative or formative evaluation, which are used to create or develop an intervention, evaluation monitoring is employed by administrators to systematically observe and track the performance of relatively established policy or program to determine if and when adjustments are required (Chen 1996; Stufflebeam 2004). This typically includes the front-line administrators of a policy or program charged with the day to day management of the intervention and/or the policy makers, funders and decision-makers who are interested in linking monitoring to planning and budgeting cycles.

Evaluation monitoring can be employed to track a variety of issues related to an intervention, including its daily operations (e.g. resource flows, number and profile of clients using services), its immediate outputs and effectives (e.g. percentage of program participants finding jobs) and the context in which it operates (e.g. policy changes, economic changes).

Evaluation monitoring encompasses a variety of methodologies and techniques. This includes: *performance monitoring*, a process which keeps track of intervention inputs, activities and immediate outputs of an intervention, as well as beneficiary satisfaction surveys and environmental scans (Bamberger et al. 2006); *results-based management*, a strong planning and management driven approach which emphasizes the clear planning and close monitoring of an intervention to achieve a discrete set of results in a context where stakeholders compete for multiple outcomes (Mayne & Rist 2006); *continuous quality improvement*, a management systems model which includes an elaborate process of establishing and measuring goals and testing new approaches to productivity through team-based experiments (Colton 1997); and, *adaptive management*, an approach that decision-makers use to help steward natural eco-systems and species in complex and unpredictable contexts (Margoluis & Safalsky 1998).

Over time, performance monitoring has become a dominant focus of evaluation activities internationally and in North America. Evaluation and monitoring is a central theme in the field of international development (Bamberger et al. 2006) and a major focus for Canadian evaluators over the last decade (Gauthier, Borys, Kischchuk, Roy, 2006). Despite – or because of– its popularity, evaluators feel that users over-use evaluation monitoring. This is because monitoring

efforts tend to track issues related to program implementation and effects, but typically examine these issues in-depth using more elaborate research methodologies. Haag and Rosengren (2001), for example, argue that Swedish civil servants have become consumed with tracking program activities and outputs of public services - inaccurate measures of their effectiveness - at the expense of evaluating the outcomes of these services.

3.2.4 Accountability

The context for evaluation shifted substantially in the mid to late 1970s. For a variety of reasons, many of the ambitious programs in the United States and other industrialized countries failed to deliver the dramatic improvements in the social and economic conditions they were designed to address. At the same time, the oil crises, high levels of public debt and two steep recessions meant fewer resources to invest in public programs. In response, many policy makers and funding organizations placed a greater emphasis on helping program administrators make do with existing programs and to hold them accountable for their actions (Haselkorn 1978).

Evaluation-for-accountability requires evaluators to assess, among other things, the extent to which interventions are meeting established objectives and targets (Washington 1965), serving or benefitting those they were designed to benefit (Hoisington 1977), implementing programs with fidelity to the original program design (Haselkorn 1978), and using resources according to agreed upon terms (Kramer & Bickel 2004). While the precise design for an accountability assessment tends to be shaped by users and the overall context, accountability focused assessments typically employ methodologies such as performance monitoring, program audits, accreditation-certificate studies, evaluation score-cards and end-of-project reports (Patton 2008).

Evaluation for accountability has become increasingly institutionalized. In 2004, the General Accounting Office of the United States Government changed its name to Government Accountability Office while central government agencies in the United Kingdom, Canada and Australia adopted a range of accountability-oriented requirements for evaluation (Bamberger et al. 2006; Patton 2008). By 2006, a majority of participants of national survey of program

evaluators in Canada reported that their evaluations were driven more by accountability requirements than for improving or judging programs (Gauthier et al. 2006).

Despite its popularity with policy makers and funders, evaluation-for-accountability evokes strong negative reactions amongst many evaluators who argue that is not a legitimate form of evaluation at all. This is because it "ignores even obscures" the causal connections between outcomes (Pressman & Wildavsky 1983, p.191), is "too narrow or is only tangential to the questions of worth" (Stufflebeam & Webster 1981; p.71) and is typically unconcerned with uncovering the processes of operations with sufficient rigor to understand and help solve problems of implementation (Perrin 1998). Stufflebeam and Webster (1981) went so far as to argue that evaluation-for-accountability was a pseudo or quasi evaluation, while Cronbach (1980) called it a "pathology" of the political system, more concerned with assigning praise or blame than understanding the intervention well enough to guide future decisions and activities.

3.3.5 Pre-Formative[9]

The fifth niche for evaluation is pre-formative evaluation. Pre-formative is distinct from the other five evaluation purposes because is employed "upstream" in the process of conceiving, designing and developing an intervention rather than "downstream" to evaluate an already developed intervention.

The discussion about using evaluation to help designers conceptualize and plan an intervention occurred relatively early in the evolution of the evaluation discipline. In the early 1970s, some observers and researchers perceived the causes of failed policies and programs to be rooted in the design − rather than implementation − of interventions and recommended that evaluators assist planners with their expertise in data and methods (Pressman & Wildavksy 1983). By 1981, Rossi and Berk argued that evaluation was an integral part of the planning of interventions as it was in the implementation and post-implementation assessment of interventions:

---

[9] This phrase is used by Patton (2010).

Evaluation research may be conducted to answer questions that arise during the formulation of policy, in the design of programs, in the improvement of programs, and in testing the efficiency and effectiveness of programs that are in place or being considered. (Rossi & Berk 1981; p. 287).

That growth of evaluation in the early design phase of evaluation has led to a number of different ways of thinking about and organizing evaluation during the design phase of programming. One of the first manifestations was "ex ante" evaluation. In contrast to an ex poste evaluation, which tracks the actual effects of an intervention once implemented, the aim of an ex ante evaluation is to estimate the possible effects of different program designs in order to help choose a final design. This is done either through a review of past evaluations on similar initiatives, quantitative modeling of interventions, the use of simulations and/or Delphi techniques. For example, evaluation researchers used behavioural models and other nonparametric estimation techniques to predict the effectiveness and optional design for a school subsidy experiment in Mexico before choosing and elaborating the eventual policy (Todd & Wolpin 2005).

"Prospective evaluation" is employed even earlier than ex ante evaluation (United States Government Accountability Office (USGAO) 1990). It involves evaluators working with the designers of an intervention to review their assumptions of causes underlying the problems they are trying to address and to surface and assess possible solutions to the problem. Moreover, prospective evaluation builds on ex ante's evaluation emphasis on prior research, modeling and simulations to include the creation and assessments of prototypes and pilot projects to yield insights that can be used for developing a larger scale program (USGAO 1990).

Chen's (1996) "practical taxonomy" contains four stages of program evaluation: program planning, initial implementation, mature implementation and outcome stage. In the program planning stage, evaluators can assist program designers with gathering background information on the rationale for the intervention (e.g. needs assessment), help develop the intervention's theory of change and logic model and then test the internal and face validity of the intervention using small pilot projects. Chen is clear to point out that evaluation in the program planning stage precedes more conventional formative evaluation in the initial program implementation stage.

Finally, Owen & Rogers (1999) created a five stage meta-model of evaluation grounded in what he argues evaluators "actually do" rather than what they "should do". The latter three stages are broadly similar to Chen's three phases, focussing on formative evaluation, performance monitoring and impact assessment. The meta-model's first two stages, however, expand upon Chen's planning stage and include a "proactive" stage which focuses on helping designers clarify what they already think they know about the problems they are trying to address and patterns of effectiveness gleaned from other evaluations. Chen includes a "claricative" phase where evaluators help program managers craft key features of the program, such as program rationale, plausible design features and realistic expectations about outcomes.

The emergence of pre-formative evaluation became widespread so quickly that it began to blur the lines between the traditional domains of planning and evaluation. As early as 1972, the celebrated evaluation-research Carol Weiss observed that evaluation was "threatening to become coterminous with policy analysis itself" (Weiss 1972, p. 34).

3.2.6 Knowledge Development

The sixth and final major niche for evaluation is to help policy makers and program designers mine and distil the evaluation findings from multiple site assessments and studies to produce knowledge that might be useful for crafting more effective interventions in the future.

As with evaluation-for-accountability, the emphasis on using evaluation to develop knowledge was not entirely new. Social scientists in the 1920s tended to treat program interventions as hypotheses to be tested and/or mechanisms to better understand the causal dynamics underlying the problem they were trying to address (The Aspen Institute 1995a). What was new in the 1980s was the emphasis on accumulating learning and evidence from multiple sites in an effort to uncover patterns of effectiveness in different domains that transcended unique contexts (Chelimsky 1997; Mark, Henry & Julnes 2000).

Since the 1980s, evaluators have undertaken a wide range of "best practice" studies, cluster evaluations, and synthesis evaluations designed to improve general understanding of social problems and identify general principles of effective intervention (Patton 2008). In the United States, for example, public agencies and foundations commissioned studies to learn the

lessons of the over 150 projects designed to better integrate services for vulnerable children and families (Schorr 1988). Internationally, organizations such as the Organization for Economic Cooperation and Development have carried out multiple studies to assess the experience of replicating micro-credit programs from lesser developed countries to Europe and North America (Snow & Buss 2001; Underwood 2006).

3.3.7 Purposes in Practice

Evaluation theorists and practitioners have surfaced, developed and elaborated on at least six major evaluation purposes over the last forty years as well as a number of secondary uses.

Together, they provide a broad continuum of evaluation activities that can be employed across a whole cycle of an intervention. This includes pre-formative evaluation to assist with conceptualizing and designing an intervention, formative evaluation to improve an intervention, summative evaluation to judge the merit or worth of an intervention, and finally, monitoring evaluation and accountability evaluation to keep track of an established intervention (Chen 1996; Owen 2004; Stufflebeam 2004). Other evaluators extend the continuum further than most by describing an evaluative process of harvesting knowledge by intervention stakeholders after it has been discontinued so that the intervention may be used in other contexts (Patton 2008; Pawson & Tilley 1997).

In practice, the traditional emphasis on formative and summative evaluation continues to be dominant concept and practice among theorists and practitioners. The authors of a popular textbook for evaluation on international development, for example, only describe summative and formative evaluation (Bamberger et al. 2006). Scriven (2004), the person credited for developing the original formative-summative distinction, recently argued there appears to be legitimate evaluation purposes beyond summative and formative, but he is uncertain what to call them, so organizes them under the broad heading of "ascriptive."

Some evaluators acknowledge the expanded scope of evaluation activities, but do so by enlarging how they define formative and summative evaluation. Stufflebeam (2004), for example, includes pre-formative evaluation activities in his Context-Input-Process-Productive model. These include "context evaluation" activities, which focus on clarifying the needs,

opportunities, goals, priorities and outcomes of a prospective intervention and "input evaluation" designed to assess alternative approaches, competing operational plans, overall feasibility and cost-effectiveness of an emerging intervention. These two stages are followed by the traditional focus on evaluating the implementation of the intervention (i.e." process evaluation") and its outcomes or impacts (i.e. "product evaluation"). However, rather than introduce a new category or purpose for evaluation, they define formative evaluation as any ex-ante activity that "guides" program designers to answer key questions in each of these four phases and summative evaluation as any ex-poste activity to "judge" the program in these phases.

Other evaluators, on the other hand, argue for a more limited role for evaluation. For example, in an apparent reversal of his earlier positions on the matter, Rossi, a well-known advocate of approaching evaluation as applied social science, argues that anything beyond formative and summative evaluation should not be considered evaluation at all:

> I do not consider designing programs and providing advice on how to manage programs to be evaluation activities. Such activities might involve the application of knowledge derived from social research, but I do not consider them social research. I do not deny that designing and managing programs are important activities calling for high levels of skill; however, they are not social research (Rossi 2004; p.129)[10]

The preferences of evaluators and evaluation users, however, do not always align. While evaluators may be eager to focus on learning and judging program effectiveness, many of the people and organizations commissioning evaluation activities are interested in performance monitoring and accountability assessments. This is particularly true in Canada were both evaluators and voluntary sector organizations report that this is the primary aim of the government agencies that fund them (e.g. Gauthier, Borys, Kishchuk, Hall, Phillips, Meilatt & Pickering 2003; Phillips & Levasseur 2004; Roy 2006; Seasons 2001).

---

[10] This statement seems to contradict his statement in 1981" Evaluation research may be conducted to answer questions that arise in the formulation of policy, in the design of programs, in the improvement of programs, and in testing the efficiency and effectiveness of programs that are in place or being considered (Rossi and Berk 1981; 287).

3.3 Major Concerns in Planning & Evaluation

Because evaluation encompasses a wide of range of disciplines and activities, is applied in a variety of contexts and settings, and is employed for so many distinct purposes, it is difficult to identify and summarize major concerns of the field. There are, however, a number of relatively consistent themes that run through discussions in professional associations and evaluation textbooks about the theory and practice of evaluation: methodological debates, roles, responsibilities and accountabilities of the evaluator, the capacity of the evaluator, and working with practical constraints, and the use and mis-use of evaluation. These challenges are described in the following sections.

3.3.1 Debates About Paradigms and Methods

The process of developing questions and designing techniques to collect and analyze, report and use data to answer those questions is a pre-occupation in evaluation. It is also an area that generates considerable debates among theorists and practitioners about the underlying paradigms on the nature of knowledge (i.e. epistemology) and being (i.e. ontology). The four major paradigms in evaluation field include the positivist, interpretivist, critical theory, scientific realism and pragmatism (Kazi 2000; Robert Woods Johnston Foundation 2008).

The positivist paradigm emerged out of the natural science tradition and assumes that: (a) there exists a reality that is distinct from the human knowledge, (b) general patterns of cause and effect human systems and behaviour world manifests general patterns of cause and effect relationships in the human world, and (c) that people can accurately describe and explain this reality. In this paradigm, the job of the evaluators is to test the effects of interventions that aim to change these patterns (i.e. working hypothesis) and judge their effectiveness against a set of pre-determined criteria. This is ideally done using experimental research designs (e.g. randomized controlled trials) designed to ensure validity (i.e. the extent to which research yields the correct answer), reliability (i.e., the ability of a method to produce the same answer whenever it is carried out), and generalizability (i.e. the extent to which the study findings can be applied outside the context of the original study).

The interpretive paradigm, often called constructivist or naturalistic, is shaped by the fields of sociology, phenomenology and anthropology. It is rooted in the idea that reality is not purely external and objective but rather is subjectively constructed by peoples' experiences, social context and interaction with others. Therefore, the task of the interpretivist evaluator is to facilitate a dialogue between intervention stakeholders about their different experiences, perspectives and judgments about an intervention, and negotiate conflicting interpretations of that reality. Interpretivist/constructivist methods, therefore, are primarily qualitative (e.g. interviews, surveys, etc.), the evaluation designs are adaptive and emergent, and the findings are captured in the form of "thick descriptions" that represent the diverse perspectives of stakeholders at a given point in time. The criteria for good interpretivist evaluation research include ethical validity (i.e. being robust about drawing out different perspective and alternative explanations) and substantive validity (e.g. being clear about the evidence underlying different perspectives (Guba & Lincoln 1989).

The critical theory paradigm emerges from the work of modern and post-modern philosophers who argue that much of reality is knowable, but its interpretation is powerfully shaped by "political, cultural, ethnic and gender-based forces" which manifest themselves into dominant social structures and everyday language (Robert Woods Johnson Foundation 2008). These structures and language skew or distort individuals' understanding of that reality. It is the evaluator's job, therefore, to shed light on the experience and effects of an intervention by pointing out and challenge the basic assumptions underlying others' interpretation of reality, which makes evaluation as much a political exercise as a research one (Mabry 2010). It may also include selectively approaching the inquiry from a discrete lens or perspective, such as feminist empiricism, which seeks to correct gender inequities by systematically interpreting reality using gender lenses and the reality of women (Greene 1997; Sielbeck-Bowen et al. 2002).

The scientific realist paradigm (sometimes referred to as the pragmatist approach) represents a middle ground between the above three paradigms (Pawson & Tilley 1997; Patton 2008; RWJF 2008). Its adherents assume that there is an objective reality but that an individual's ability to know and understand it is either imperfect (e.g. critical realists) or only knowable from the individual's unique perspective (e.g. subtle realists) (RWJF 2008). They further believe that while reality is objective, its cause and effect patterns are highly contextual – rather than

universal – which makes it difficult to generalize findings from one context to another (Pawson & Tilley 1997). Rather than ask "does the intervention work? scientific realists seek to understand "what works for whom, why and in what contexts?" Their preferred approach to answering these questions is to employ mixed methods, qualitative and quantitative data, careful and purposeful sampling, fixed and emergent designs, and weave together facilitate dialogue amongst stakeholders, peer reviews, and literature to help make sense of evaluative data (RWJF 2008).

As with any debate among paradigms, it is difficult to find common ground between people operating with such different perspectives. Up until the 1960s and early 1970s, the positivist paradigm was dominant in evaluation theory and practice and shaped the majority of government and philanthropic efforts by striving to "scientifically investigate" the causes and effects of larger scale policies and programs to reduce poverty and renew distressed neighbourhoods (O'Conner 1995; Patton 2008; Rossi 2004). When it became clear that positivist evaluations by themselves not always feasible nor able to yield the type of data and knowledge required by evaluation users, evaluators and evaluation experimented and elaborated on interpretivist, scientific realist and critical theory approaches (Guba et al. 1989).

The Canadian Evaluation Society, the American Evaluation Association, and Joint Committee on Standards for Educational Evaluation strive to create a "big methodological tent". The emphasis of both organizations to systematic research that embraces mixed methods appears to reflect the paradigm of scientific realism. The support for this inclusive approach is uneven. The agencies of the United States government and World Bank, for example, have both recently formerly confirmed their preference for quantitative methods such as randomized controlled trials (Bamberger et al. 2006; Patton 2008).

3.3.2 Uncertain Roles, Relationships & Accountability

Evaluators have a broad agreement about their primary roles, relationships and accountability with evaluation users. The primary responsibility is to help policy makers, planners and administrators make decisions about their interventions based on the systematic gathering and analyzing of data. The manner in which they accomplish this is captured in

professional guidelines of professional associations across the world. The members of the Canadian Evaluation Society, for example, have agreed that evaluators are meant to be accountable for their performance and the results of their work, including:

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.

Beyond these broad accountabilities and responsibilities, the precise roles of the evaluators, and relationships to evaluation users involved in fulfilling these obligations is less clear. To some degree, these roles vary with the purpose of the evaluation. As a rule of thumb, evaluators in summative evaluations which aimed at judging the merit or worth of a program or assessing the extent to which program administrators are accountable in implementing the intervention, operate at arms-length from administrators and treat funders and decision-makers as their primary intended users. Evaluators involved in pre-formative, formative and monitoring evaluations, on the other hand, tend to work more closely with – and are accountable to – the administrators and persons involved in the daily operations of an intervention so that they can provide them information designed to improve the design or implementation of the initiative.

The picture becomes murkier when evaluators who prefer to operate from a particular paradigm of social inquiry, each with a different perspective on the roles of evaluators and their relationships to the intended users, enter the picture. Positivist evaluators argue that their job is to act as an external, dispassionate and arm's-length specialists whose job is to deliver data findings and judger merit or worth against pre-set criteria (Scriven 2004). Interpretivist evaluators, on the other hand, see their role as assisting intervention stakeholders to surface and negotiate different experiences and perspective (Guba et al. 1989). The role of the evaluator who operates from a

critical theory perspective is to help primary intended users and other stakeholders deconstruct their own artificial interpretation of the world and advocating for the participation and insight of diverse perspectives,. This is a role that may create tension with primary intended users (Fetterman 1996; House 1990; RWJF 2008). Finally, the scientific realist tends to look to the evaluation purpose to shape their roles but is open to fulfill whatever role is most helpful in a given context or situation (Pawson et al. 1997).

A final level of complexity on the question of roles and relationships is added when evaluators seek to address other issues in the evaluation. Evaluators concerned about improving the likelihood that evaluation findings will be used, for example, argue that the evaluator has the extra role of actively interacting with primary intended users in the design, implementation and use of evaluation activities and findings, and tailoring the process in a way that meets their idiosyncratic preferences, while meeting the demands of good evaluation practice (Alkin 2004; Patton 2008; Stake 1974). Similarly, evaluators and evaluation users who are eager to extend participation beyond direct decision-makers to include those directly and indirectly affected by the intervention need are more apt to operate as a member of the evaluation team and thus require strong facilitation skills (House 1990; King 1998; Weaver & Cousins 2004).

The contingent nature of roles, responsibilities and accountabilities reveals a broader pattern of debate and disagreement in the evaluation community about narrow and broad definitions of evaluation. Advocates of a narrow definition, typically steeped in the social science tradition of evaluation, warn that evaluators who venture into activities beyond gathering data in order to judge merit or worth, risk performing the role of organizational development consultants (House 2004; Stufflebeam 1967). Evaluators who perceive evaluation as one part of a broader approach to organizational and policy learning, on the other hand, are comfortable with evaluators playing whatever roles are necessary to encourage a culture of learning and evidence-based decision-making (King 1998; Owen & Rogers 1999; Preskill & Torres 2001).

3.3.3 Robust Design in the Face of Constraints

The third challenge refers to dealing with practical constraints to design an evaluation design that is able to deliver an appropriately robust level of data and analysis to answer the key

evaluation questions of its intended users. Regardless of the epistemological preferences of evaluators and evaluation users, evaluators almost always face a number of logistical constraints.

The appropriate "burden of proof" varies case by case, but it is in part shaped by the purpose of the evaluation and whether the stakes are low, medium or high. The stakes and burden of proof is high, for example, in summative evaluation when deciding whether a program should be terminated, continued or scaled up. On the other hand, the stakes and burden of proof are comparatively lower in formative or regularly performance monitoring (Patton 2008).

One of the most common challenges is working with limited resources. Many organizations interested in evaluation findings do not invest sufficient funds to carry out an effective evaluation (Alkin 2004; Bamberger et al. 2006; Rossi et al. 1999; Seasons 2002; Patton 2008). Evaluators and evaluation users, for example, have identified limited and shrinking resources as a major barrier to employing consistently good evaluative practices across in territorial, provincial and federal government agencies and voluntary organizations in Canada (Gauthier et al. 2009; Seasons 2001). This pattern appears to be repeated in the non-profit and voluntary sector (Hall, Phillips, Meilatt & Pickering 2003; Phillips & Levasseur 2004).

Limited time is another major constraint. Evaluators are often called in to provide support after an intervention has already started, which does not allow them to prepare the intervention for effective evaluation, carry out pre-tests of instruments, and/or develop a strong baseline required for the longitudinal tracking of data (Bamberger et al. 2006). They may also be asked to take on a large data collection effort within a relatively short period of time (Rossi et al. 1999). Finally, evaluators are often pressured to present solid evaluation findings well before the intervention has managed to work itself through. In her historical review of efforts to evaluate neighbourhood renewal programs, for example, O'Conner (1994) describes how the short term political cycles and constant changes of federal policy priorities made it virtually impossible to carry out the kind of longitudinal research required to determine if there were any measureable and durable effects of suddenly appearing and disappearing programs.

The final – and some argue endemic – logistical challenge is working in situations where critical data is missing or difficult to collect. The number and variety and number of constraints

on data is great and include, among others, lacking access to baseline data, difficulty in finding comparison data with other subjects, sample sizes that are too small to be useful, and challenges collecting data on sensitive topics and/or from difficult subjects (Alkin 2004; Bamberger et al. 2006; Ely et al. 1991; Rossi et al. 1999).

Evaluators can identify and work around constraints by employing a systematic approach to exploration, decision-making and design. This includes evaluability assessments to assess the extent to which an intervention is ready for an assessment, such as whether it has clear goals, coherent conceptualization, measures of success (Smith 1989; Wholey et al. 1994); readiness assessments to determine the ability of organizations or communities capacity to participate in and use evaluative data (Kusek & Rist 2002); an evaluation scope of work to flush out and flesh out the broad parameters required to do begin practical design, such as determining primary users, purposes of the evaluation, evaluation questions, data sources, budget, etc. (United States Agency for International Development 1996); and rapid-feedback assessments, which focuses on easily collected information and offers the opportunity to test methods and the extent to which evaluative data are useful and will be used (Wholey 1983).

Ideally, exploratory and scoping techniques can assist evaluators and evaluation users develop a strong enough understanding of the intervention, its context, the expectations of users and stakeholders, and the evaluation constraints to prepare a customized and "good enough" evaluation design. It may also prompt them to decide that an evaluation should not take place because it is not possible to develop and carry out a sufficiently robust evaluation, and/or the possibility that the findings are unlikely to be used for intended purposes.

In practice, evaluators often do not have the time, resources or permission to employ such a systematic approach and end up simply doing their best to spot and overcome these constraints as they go (Bamberger et al. 2006).

3.3.4 Perspectives, Priorities, Politics and Power

Most evaluations have diverse stakeholders with their unique – sometimes conflicting – values, interests, perspectives and priorities that they would like to see reflected in the evaluation design. For the evaluation of a program, factors that can affect the success of an evaluation

process can include the views of those with overall decision-making authority, program administrators and staff, funders, political advocates, beneficiaries as well as other organizations (e.g. program partners) and individuals (e.g. residents of the neighbourhood in which a program is located) (Cousins 2004).

Stakeholder preferences influence all aspects of an evaluation process, from beginning to end, including: (a) who is selected as the evaluator, (b) selection of stakeholders, (c) clarification of goals, (d) priority evaluation questions, (e) design and instruments, (f) indicators and instruments, (g) choice of stakeholders, (h) role of evaluators, (i) budget and time, (j) choice of audiences for the evaluation, and (k) timing, packaging and strategy for disseminating evaluation findings (Bamberger et al. 2006).

The process of surfacing and negotiating these preferences can play out in small and nuanced or large and formal ways. When asked for their input on evaluation instruments, for example, the managers of programs often prefer indicators and sources of information that shed their efforts in a positive light (Willcocks & Lester 1999) and habitually favour reliable and familiar data over data that may shed deep insight into the effectiveness of their work (Martin 2007). On a larger scale, O'Conner (1994) and Weiss (1991) conclude that political and institutional dynamics have been the largest barriers to effective evaluations of neighbourhood renewal initiatives and other large government sponsored programs.

Stakeholders with more power have a greater influence in shaping evaluations. Program managers and funders, in particular, have more direct and frequent contact with evaluation process, resources and formal authorities than other stakeholders (House 1990). Evaluators themselves are apt to pay comparatively more attention to their concerns, motivated in part by a desire to improve the probabilities that their evaluation will be used and/or by a reluctance to "bite the hand that feeds them." (Bamberger et al. 2006, p. 118)

Evaluators can use three broad strategies to navigate and address the politics and power of evaluation. The first is to surface, make explicit and attempt to negotiate the varied preferences of stakeholders in the exploratory phase of an evaluation (Bamberger et al. 2006; House 2004; Wholey et al. 1994). This can be facilitated through the use of aides such as

utilization-focused checklist (Patton 2008), evaluation scopes of work (USAID 1993) and a variety of other "exploratory" techniques (Owen 2004).

They can also address multiple perspectives in the design of the evaluation, through methods such as "goal-free" evaluation, which encourages evaluators to explore the unintended effects of interventions beyond those intended by program designers (Scriven 1972), adversarial evaluation, which facilitates active debate between different perspectives (Wolf 1975), and "decision-theoretic" assessment which addressed multiple outcomes "explicitly valued" by multiple stakeholders (Pressman & Wildavsky 1983).

Finally, evaluators can play an active role in (re) balancing power by engaging a broad number of users and stakeholders and helping them negotiate different values, perspectives and judgements. This approach is manifested in different ways and includes participatory evaluation (Weaver et al. 2004), empowerment evaluation (Fetterman 2001), deliberative democratic evaluation (House & Howe 1999), feminist and advocacy evaluation (Green 1997) and naturalistic evaluation (Lincoln & Guba 1985). These processes require evaluators to have additional skills in facilitation, negotiation, conflict resolution, power analysis, and an ability to "speak truth to power" (Fetterman, Kaftarian & Wandersman 1996; Weaver et al. 2004).

3.3.5 Uneven Evaluator Capacity

Another major issue in the field of evaluation is ensuring that evaluators have the core competencies required to design and deliver effective evaluations. This is an endemic conversation is the evaluation community. For example, there were over forty panels, debates, roundtable and think tanks on evaluator competencies at the American Evaluation Associations from 2003-2006 (Dewey, Montrose, Schroeter, Sullins & Mattox, 2008).

Evaluators who disagree about paradigms and evaluators roles have made it difficult to develop a common and broad agreement among evaluators about the core competencies of the evaluation (Bamberger et al. 2006). They also have difficulty in agreeing on the relative importance of competencies. For example, a team of researchers found that educational institutions tended to rate competencies in research methods as the most important skills in

evaluators while employers and evaluation users tended to emphasize interpersonal skills, project management capacity and communication (Dewey et al. 2008).

In some cases, professional evaluation associations have developed formal consensus about the competencies that evaluators require to be effective. The Canadian Evaluation Society (CES), for example, is the first association in the world to agree upon a list of competencies organized around five broad themes (King, Ghere & Minnema 2005):

1. Reflective Practice competencies that focus on the fundamental norms and values underlying evaluation practice and awareness of one's evaluation expertise and needs for growth.

2. Technical Practice competencies that focus on the specialized aspects of evaluation, such as design, data collection, analysis, interpretation and reporting.

3. Situational Practice competencies that focus on the application of evaluative thinking in analyzing and attending to the unique interests, issues, and contextual circumstances in which evaluation skills are being applied.

4. Management Practice competencies that focus on the process of managing a project/evaluation, such as budgeting, coordinating resources and supervising.

5. Interpersonal Practice competencies focus on people skills, such as communication, negotiation, conflict resolution, collaboration, and diversity.

The Canadian Evaluation Society offers a variety of training programs organized around these competencies as well as the opportunity to receive a professional designation after demonstrating their capacity in these areas.[11] Their efforts complement a variety of capacity building opportunities offered through universities, other professional associations, intermediaries (e.g. the Evaluator's Institute) and communities of practice (e.g. Evaluation Wiki, Eval Talk).

While people in the field are becoming clearer about what they feel constitutes good evaluator capacity, there appears to be a gap between the supply and demand of capable evaluators (Gauthier et al. 2009; Seasons 2001). Limited dollars devoted to evaluation and a gradual shift to accountability-based evaluation appears to have prompted some evaluators to leave the field and discouraged others from joining (Gauthier et al. 2009).

---

[11] See the Canadian Evaluation Website at: http://www.evaluationcanada.ca/site.cgi?s=5&ss=10&_lang=en

3.3.6 (Mis) Use

A final major theme in the evaluation field is the challenge of getting would-be users of evaluation findings to actually use the findings. As early as the 1960s, researchers reported a dearth of examples where evaluation informed policy (Williams & Evans 1969), a "failure" to use evaluation (Wholey et al. 1970), and that the influence of social science research and evaluation on program decisions was "with few exceptions, nil" (Deitchman 1976). In 1975, a celebrated educational researcher published an article with the provocative title "Evaluation: Who Needs It? Who Cares?" (Alkin 1975).

The growth and sophistication of the evaluation field does not seem to have changed the overall pattern of use a great deal. In 2005, the Evaluation Gap Working Group found few rigorous evaluation studies of the more than thirty four billion dollars spent on foreign assistance that year (Patton 2008). Landry (2003) and his colleagues found few government officials made regular use of social science research and evaluation. Reminiscent of Alkin's 'who cares' article thirty years earlier, two senior administrators wrote "Why Measure? Nonprofits Use Metrics to Show That They are Efficient, But What if Donors Don't Care?" (Cunningham & Ricks 2004).

Evaluators have identified a variety of reasons for low use. Some reasons related to the subject of the evaluation and the evaluation itself and included (a) poor intervention design, (b) limited evaluation expertise, (c) weak methodology because of insufficient resources, (d) poor timing, (e) inappropriate dissemination and communication of results, and (f) evaluators not trusted or credible in the eyes of the evaluation users. Others factors related to political issues such as data that was irrelevant – or even in conflict with – the opinions, values and interests of decision-makers (Carden 2009; Pressman & Wildavsky 1983; Patton 2008).

Evaluators such as Burry (1984), Johnson (1998) and Patton (2008) concluded that the factor that most influenced the use of evaluation findings was the personal factor - that is, the presence of identifiable people who cared about the evaluation and its findings and were committed to using them for decisions. They further concluded that evaluators who were interested in improving the uptake of evaluation findings, could improve the probabilities of this happening by working closely with primary intended users throughout the evaluative process so

that they could (a) help shape the evaluation design to reflect their idiosyncratic needs and preferences (b) have an additional level of ownership of the process and findings. Weiss, one of the first persons to become concerned with issues of use and one of the field's longest standing researchers on the topic, concludes that research quality is the greatest predictor of use (Weiss & Bucuvalas 1980; Weiss 1972, 1991).

The concern for use in evaluation has created a place for a special branch of evaluation. In a review of major evaluation theorists in North America, Alkin (2004) concluded that Stufflebeam, Wholey, Alkin, Owen, Cousins, Fetterman, Preskill, Patton and King are all primarily concerned with designing evaluations that will be used. By the early 1990s, Wholey, Hatry & Newcomer (1994) declared that utilization had become the watchword of the evaluation profession. It has also led to a broader understanding of use. While most evaluators are concerned with "instrumental" or direct use of findings, evaluators also acknowledge "process use" where the process of evaluation leads to changes the way evaluation users think about the challenges they are trying to address and/or strengthen the evaluative culture and capacity of the evaluation users (Patton 2008).

Some evaluators feel that the emphasis on utilization has gone too far. Patton (2008) recently declared that his major concern was the danger of "mis-use". These include: *political uses,* when evaluation processes and findings are used to selectively support a political position, to legitimize an already established political decision; *mechanical use or compliance use* refers to the action of going through the motions of evaluation without intention of using the findings to shape decisions; and, *mischievous uses* refer to the deliberate suppression, misrepresentation, or unbalanced use of evaluation findings to influence opinions and decisions (Patton 2008: pp. 112-113).

While evaluators are unlikely to design an evaluation with mis-use in mind, they may nonetheless feel pressure to consider these uses by evaluation users, and they have little control on how their evaluation results are used once their work is completed. Ultimately, the consensus seems that even when evaluators do "everything right", all they can do is improve the probability – rather than guarantee – that evaluation findings will be used.

3.4 Summary & Conclusion

This chapter has described the link between planning and evaluation, explored in more depth the different purposes for which evaluation activities can be employed, and reviewed some of the major issues related to the theory and practice of evaluation.

Five out of six evaluation purposes can be described as the systematic application of data and knowledge to evaluate the downstream design, operations, effects and new knowledge of an intervention(s): i.e. summative, formative, monitoring, knowledge development and accountability. Pre-formative evaluation, in contrast, involves the application of data and critical thinking to the "upstream" work of conceptualizing and designing an intervention. Among these major purposes, formative and summative evaluation continues to dominate the thinking and practice of the evaluation field.

Above and beyond these core evaluation purposes, the theory and practice of evaluation is shaped by a number of debates and challenges. These include debates about methodology, clarifying appropriate roles, responsibilities and accountabilities, designing effective evaluations amidst multiple constraints; working with political considerations; improving the use and avoiding the mis-use of evaluation; and issues related to the capacity of evaluators.

Chapter Four: Developmental Evaluation

4.0 Introduction

This chapter describes the new concept and approach of developmental evaluation prepared by Michael Quinn Patton. It includes a description of the emergence of developmental evaluation, its primary purposes and "niches" for use, roles, responsibilities, capacities and accountability of evaluators, approach to methods, challenges, and emerging criticisms. The chapter ends with a number of emerging questions about how developmental evaluation is understood and applied by evaluators beyond Michael Quinn Patton.

4.1 Genesis

Michael Quinn Patton typically uses the following story to describe the moment he used the term developmental evaluation.

When Patton told the managers of a leadership program in Minnesota that after several years of developing the program and formative evaluation activities that they needed to stop adapting the model so they could prepare it for a summative evaluation, they expressed disbelief. They responded that that they did not want to stabilize the program because they needed to continually adapt the model to reflect the realities of new demographics, shifts in funding, emerging technologies, and the need to move the program from an urban to rural area. Moreover, while they were flattered that other communities were interested in adopting their model, they were not so interested that they were willing to subject it to a summative evaluation if that meant they could no longer tinker with it (Patton 2010).

Patton replied that while he understood eagerness to be responsive to a dynamic environment, he was in fact hired to carry out an evaluation of the program. This conventionally involved first employing a formative assessment to work out the bugs of a new program and then a summative evaluation to assess its longitudinal effects and worth.

Clearly frustrated, the Program Director asked Patton, "Formative and summative evaluation, is that all you evaluators have to offer?" Thinking quickly on his feet, Patton spontaneously and awkwardly offered the possibility of doing a "developmental evaluation" where evaluative thinking was used to help a group continually develop and adapt a program. Satisfied, the Director replied that was precisely the kind of evaluation in which they were interested in, and the group and Patton used evaluative thinking and methods to inform the program managers' thinking and decision-making process for the next several years without any intention of stabilizing or judging the program model.

Since that time, Patton has described his emerging thinking and practice on developmental evaluation in a variety of publications (1994, 1999, 2006 and 2008) and published a book devoted specifically to the topic in 2010.

4.2 Purpose & Niches

Patton describes developmental evaluation as bringing evaluative thinking and data to a group's effort to conceptualize, develop, test – and when necessary, to significantly adapt – an intervention.[12]

> Developmental evaluation refers to long-term, partnering relationships between evaluators and those engaged in innovative initiatives and development. Developmental evaluation processes include asking evaluative questions and gathering information to provide feedback and support developmental decision-making and course corrections along the emergent path (Patton 2008, p.30).

He argues that this purpose is distinct from the more traditional purposes of formative evaluation, which aims to help administrators improve an intervention, and summative evaluation, which seeks to help the ultimate decision-makers judge its overall merit or worth. The aim of developmental evaluation is to assist those developing an intervention:

> Developmental evaluation focuses on developmental questions: What's being developed? How is what's being developed and what's emerging to be judged? Given what's been developed so far and what has emerged, what's next?

---

[12] For the remainder of this paper, intervention will be used to represent a policy, program, service or project that is being created or adapted by policy makers, program designers or front line staff.

The developmental evaluator inquires into *developments*, tracks *developments*, facilitates interpretation of *developments* and their significance, and engages with innovators, change agents, program staff, participants in the process, and funders around making judgments about what is being developed, how it is being developed, the consequences and impacts of what has been developed, and the next stages of development (Patton 2010, p. 227).

Patton describes the process of development as typically iterative, emergent and messy. Those in charge of developing or adapting an intervention often find it difficult to agree to define the problem they would like to address or the change they would like to make. They struggle to decide which avenues to pursue and which ones to reject. Their early probes reveal new things about the cause and effect dynamics underlying the issue they are trying to address, prompting them to rethink their definition of the problem and develop new options. They must suddenly radically redesign what have to this point seemed to be promising – sometimes even proven -- interventions in the face of fast moving, ever-changing contexts (Patton 2006, 2008, 2010).

Developmental evaluation aims to assist those immersed in the complex process of development to employ a systematic approach to critical thinking and gathering and using data to inform their thinking and decisions in that process (Patton 1994, 1999, 2006, 2008). This requires that evaluators must be careful not to indiscriminately approach developmental situations with same orientation and roles they employ in more traditional summative and formative evaluation which depend on the existence of clear goals and a tangible – albeit still evolving – intervention.

Patton describes developmental evaluation as a "small and demanding niche" and identifies five broad situations in which it may be appropriate (Patton 2008, 2010). These are:

1. *Pre-formative development of a potentially scalable innovation* to the point where it is ready for traditional formative and summative evaluation. Pre-formative developmental evaluation works with emerging ideas and visionary hopes in a period of exploration to shape them into a potential model that is a more fully conceptualized, potentially scalable intervention within the framework of the adaptive cycle. As models emerge out of exploratory and innovative initiatives, some may move into more traditional formative and summative evaluation to determine scalability and generalizability, while others remain in developmental mode, either undergoing further development or continuous experimentation in the search for new models.

2. *Ongoing development* in adapting a project, program, strategy, policy, or other innovative initiative, to new conditions in complex dynamic systems.

3. *Adapting effective general principles to a new context* as ideas and innovations are taken from elsewhere and developed within a new setting, the work of developmental evaluation in the dynamic middle between top-down and bottom-up forces of change.

4. *Developing a rapid response* in the face of a sudden major change or a crisis, like a natural disaster or financial melt-down, exploring real time solutions and generating innovative and helpful interventions for those in need.

5. *Major systems change and cross-scale developmental evaluation*, providing feedback about how major systems change is unfolding, evidence of emergent tipping points, and/or how an innovation is or may need to be changed and adapted as it is taken to scale. Principles are shared and disseminated in an effort to have broader impact. Horizontal scaling across systems or vertical scaling to broader systems may involve more than adaptation. These dissemination and scaling processes can evolve an essentially new development, the emergence of which can be documented and analyzed as part of a developmental evaluation (Patton 2010, pp. 309-13).

These niches expand upon the pre-formative niche described in the previous chapter. While the evaluators have employed pre-formative evaluations to assist in the conception and design of the intervention since the 1970s, Patton argues that the dynamics of conceptualizing, designing and adapting an intervention are broadly similar in the other niches as well (See Table 4-1).

Patton introduced these five niches for developmental evaluation gradually. He first described the niche of the ongoing adaptation of an intervention in 1994. In 2005, he explored the dynamics of social innovation and complexity and implications in the book Getting to Maybe (Westley et al. 2006). In 2008, he introduced the niches of cross-scale complexity and in 2010 he included the niche of replication and situations of crises or major change. He also confirms that other niches for developmental evaluation may emerge (Patton 2010).

Table 4-1 Summary of Niches for Developmental Evaluation

PRIMARY DEVELOPMENTAL EVALUATION PURPOSE	COMPLEX SYSTEMS CHALLENGES	PRIMARY SPECIFIC DEVELOPMENTAL EVALUATION USES	IMPLICATIONS
1. Ongoing development	A project, program, policy, or other innovative intervention is being implemented in a complex dynamic environment.	Adapt to changing social, political economic, environmental, technological, and demographic patterns.	The program doesn't intend to become a fixed, standardized model. It does, however, identify effective principles that inform its ongoing development.
2. Adapt effective principles from elsewhere to a new context (which can be local, regional, national, or cross-national).	An innovative project, program, policy, intervention, or idea is being disseminated by social innovators; people in a new area are interested in developing their own version based on adaption of effective principles and knowledge from elsewhere.	Identify relevant principles, knowledge, and ideas to be adapted; help keep the adapters attentive to larger and broader guiding principles, knowledge, and ideas; document the consequences of adaptations of and departures from what has been done elsewhere.	Developmental evaluation operates in the middle between top-down and bottom-up forces for change, facilitating synthesis of top-down forces (general principles and knowledge being disseminated) and bottom-up sensitivity to context, experiences, capabilities, and priorities.
3. In the face of a sudden major change (a black swan event) or a crisis, exploring real time solutions and generating innovative responses.	In the midst of crisis, there is no time for formal model development. Action is needed now, but what to do is uncertain and contentious. Rapid feedback is needed about efforts to intervene to mitigate the crisis or disaster.	Support development of new initiatives to meet emergent needs and crisis conditions; facilitate creative collaboration of local, national, and/or international response teams and innovators by bringing evaluative thinking into rapid response initiatives and crisis management.	Planning, execution, and evaluation occur simultaneously. Everything must happen at once. The stakes can be quite high. Errors and miscalculations need to be corrected quickly. Decision makers need the best information and analysis available even as they have to make decisions with incomplete data. Time is of the essence. Credible, relevant, and real time data can save lives.
4. Pre-formative development of a potentially broad-impact, scalable innovation.	Changing and dynamic systems require innovative solutions and creative new approaches to worsening conditions. Social innovators aspire to major change with broad impact, expecting to engage in disruptive systems change with a new model. But *that new model does not exist and needs to be developed*, reorganizing and exploring new possibilities as old systems show signs of collapse and dysfunction (and may have already fallen apart).	Helping the innovators track their evolving understanding of the problem and their response, creating manageable and testable boundaries around the innovation; Support getting a potential new model sufficiently well-developed and formulated that it can be further developed through formative and then summative evaluation, to identify whether it is ready to be taken to scale (broadly disseminated) for major impact.	As models emerge out of exploratory and innovative initiatives, some may move into more traditional formative and summative evaluation to determine scalability and generalizability, while others remain in developmental mode, either undergoing further development or continuous experimentation in the search for new models.
5. Major systems change and cross-scale developmental evaluation	An innovative intervention has been developed, then formatively evaluated, and successfully summatively evaluated. The success is sufficient that social innovators and funders now want to take the innovation to scale, expanding to new systems horizontally (more of the same units elsewhere, e.g., new cities) as well as vertically (from cities to regions and entire countries).	Look out for system change indicators and any emergent tipping point; gather feedback on how a model is and needs to be adapted as it is taken to scale, including the possibility that either horizontal or vertical scaling will constitute not just an adaptation but essentially evolution of a new development.	Adaptive cross-scale innovations assume that the complex nonlinear dynamics and adaptive cycles of scale will require agility, responsiveness, and adjustments. DE provides the data to be agile, responsive, and adaptive in the face of cross-scale dynamics.

Adapted from Patton 2010, pp. 308-313.

Developmental evaluation can be carried out concurrently with other types of evaluation, a situation Patton (2010) calls "patch evaluation". For example, a group of organizations working overseeing the implementation of a ten year plan to end homelessness may use formative evaluation to inform in the early days of operating a new homeless shelter, a summative evaluation to help determine whether a pilot program of wrap around services for homeless persons should be sustained, discontinued or scaled up, and a developmental evaluation to help guide the ongoing evolution of their overall strategy in the face of new learning's, stakeholders and shifting contexts.[13]

4.3 Roles & Relationships

In developmental evaluation, the evaluator operates as a member of the team that is creating and adapting an intervention(s). This is a long-term partnership and process in which:

> The evaluator is part of a team whose members collaborate to conceptualize, design and test new approaches in a long-term, on-going process of continuous improvement, adaptation and intentional change. The evaluator's primary function in the team is to elucidate team discussion with evaluative questions, data and logic, and to facilities data-based assessments and decision making in the unfolding and developmental processes of innovation (Patton 2008, p. 30).

The developmental evaluator may be a current member of the group or someone from outside the organization. Patton argues that the external-internal distinction is not a primary concern in developmental evaluation:

> Because of this long-standing differentiation between the roles of external versus internal evaluators, one of the first questions I get in presentations and training sessions is whether the developmental evaluator should be internal or external. I respond that developmental evaluation is a role not a location. The developmental evaluator supports development. I have conducted developmental evaluation as an internal evaluator and an external evaluator, and know of both internal and external evaluators who have played the role of developmental evaluator. In either case, the evaluator becomes part of the development process, asking

[13] An example used in 2005-2006 during the Du Pont-McConnell sponsored training on developmental evaluation.

developmental evaluation questions, bringing evaluative thinking to the innovation team, and supporting ongoing decision-making, adaptations, and development with real time data and feedback. The first-order identity, then, is not that one is an internal or external evaluator, but that one is a developmental evaluator (Patton 2010, p.65).

While Patton does feel the internal-external distinction is a secondary concern for an evaluator, it does appear to have practical implications. Several of the ten participants in the Du-Pont-McGill training group in developmental evaluation who operated as an internal team members reported that they found it difficult to maintain their role as an evaluator – rather than just another member of the team able to fully participate in all aspects of the development work. Evaluators from outside of the organization, on the other hand, complained that they often felt they missed the opportunity to bring evaluative thinking to bear in the many spontaneous developmental moments that emerged when they were off site. [14]

4.4 Methods

Patton consistently points out that effective evaluation in developmental situations does not depend on a specific set of methods or techniques. He firmly roots the approach in the paradigm of utilization-focused evaluation which is emphasized methodologically and requires the evaluator to tailor the evaluation design to the unique circumstances of the evaluation situation and evaluation users.

> Developmental evaluation isn't some particular methods of recipe-like steps to follow. It doesn't offer a template of standard questions. It's a mindset of inquiry into how to bring data to bear on what's unfolding so as to guide and develop the unfolding. What that means and the timing of the inquiry will depend on the situation, context, people involved, and the fundamental principle of doing what makes sense for program development (Patton 2010: pp.75-6).

He has gone out of his way to describe how developmental evaluation can employ methods typically used for summative evaluations. This includes an account of how Barack Obama's Presidential election team used experimental and comparison designs to quickly test and refine different campaign messages delivered through an elaborate internet-based communication strategy. Once they had an idea of the likely effect of different messages, they

---

[14] This is based on the author's participation in all three training sessions in 2005-2006.

chose the ones most likely to have the desired effect and therefore disseminated across the country. This was in done as part of a larger research methodology which used a variety of other qualitative and quantitative methods as well (e.g. focus groups, key informant interviews, content analysis of media, etc.).[15]

In keeping with the emphasis on utilization, and reflecting the emergent nature of developmental situations, Patton argues that methods in developmental evaluation should be flexible and adapted as the conditions for the evaluation evolve. In fact, he goes to great lengths to argue against developing "off the shelf" frameworks or models and the pursuit of perfect designs:

> [I]t's worth emphasizing that no definitive list of developmental evaluation inquiry approaches can or should be constructed. Developmental evaluation creatively adapts whatever approaches and methods fit the complexities of the situation and are responsive, appropriate, and credible to social innovators in opening up new understandings and guiding further development. In being creative, the developmental evaluator is also practical and pragmatic, doing the best job possible within available resources and other constraints. Constraints always exist and do what constraints do -- constrain. Our ability to think of alternatives is limited. Resources are always limited. Time is of the essence. We do what we can. Part of what we can do is adapt other inquiry traditions to the purposes of developmental evaluation.

Finally, as illustrated by the following quote, Patton repeatedly confirms the need for adaptive evaluation design and methods:

> Dynamic complexities don't slow down or wait for evaluators to write their reports, get them carefully edited, and then approved by higher authorities. Any method can be used but will have to be adapted to the necessities of speed, real-time reporting and just-in-time, in-the-moment decision-making. That is a major reason the developmental evaluator is part of the innovation team, to be present in real time as issues arise and decisions have to be made […] Contrary to the usual practice in evaluation of fixed designs that are implemented as planned, developmental evaluation designs can change as the innovation unfolds and changes (Patton 2010, p. 335-6).

Though keen to avoid one-size-fits-all approach to method, Patton advocates for the use of inquiry frameworks that reflect the nature of different developmental situations. He provides

---

[15] Patton used this example in his presentation on developmental evaluation at the Canadian Evaluation Society conference in Ottawa in May 2009.

examples of ten unique inquiry frameworks and that were useful in previous evaluations. These include: different sets of questions (e.g. descriptive, values-driven questions, wicked) triangulated learning framework, values driven inquiry, appreciate inquiry, and complexity-based inquiry (Patton 2010).

4.5 Situational Recognition

Developmental evaluation is a small and demanding niche that is distinct from other types of evaluation (Patton 1998, 2010). In his later writings, Patton describes the importance of "situational recognition" to determine the extent to which a situation is suitable for developmental evaluation or any other type of assessment.

This process can be aided by sensitizing concepts such as (a) frameworks which distinguish between simple, complicated, complex and chaotic contexts, each of which require orientation and approach to evaluation, and (b) the panarchy framework, which has four phases of the eco-cycle of an intervention, including an exploratory or pre-formative phase which is suitable for developmental evaluation.

These frameworks can be used in a variety of ways, including in informal conversations with evaluation users, self-assessments and as a process for surfacing questions that might be pursued in the evaluation.

4.6 Accountability

Developmental evaluation requires a unique approach to accountability. The accountability of an evaluator in formative evaluations is usually to administrators to help them improve their model based on a rigorous process of gathering and analyzing data on implementation. In summative evaluation, it is decision-makers – typically funders, policy makers, and executives – who are active in summative evaluation for judging merit or worth.

In developmental evaluation, the innovators are ideally accountable to themselves and have a high degree of commitment to results, data-based decision-making and learning,

characteristics that Patton and his colleagues argue is common among social innovators (Westley et al. 2006):

> Complexity-based developmental evaluation shifts the locus and focus of accountability. […] for vision-and-values-driven social innovators the highest form of accountability is internal. Are we walking the talk? Are we being true to our vision? Are we dealing with reality? Are we connecting the dots between here-and-now reality and our vision? And how do we know? What are we observing that's different, that's emerging? These become internalized questions, ferociously, continuously, because they want to know (Patton 2010, p. 13)

Developmental evaluators recognize that in order to work adaptively, innovators will benefit from the organizations and people who underwrite their work and who give them the room to experiment and operate:

> Those funding innovations join in the questioning and need to understand that the seriousness and resulting learning constitutes accountability (Patton 2010, p. 14).

This understanding and support is sometimes in short supply. In his co-exploration of the dynamics of social innovation, Patton (Westley et al. 2006) describes how funders that expect or require replicable models, guaranteed results, and development on schedule can inadvertently short-circuit the developmental process and, by extension, the effectiveness of developmental evaluation.

4.7 Evaluator Capacity

Patton argues that developmental evaluator require the same competencies required in more traditional forms of evaluation (e.g. research expertise, strong communication skills, etc.).[16]

He further argues that the "small but demanding niche" of developmental evaluation requires some unique capabilities. This includes (a) knowledge about relevant patterns of effectiveness gleaned from other evaluation assignments and research in the field (Patton 1999), (b) interpersonal and facilitation skills (Patton 2008), and (c) comfort with ambiguity and uncertainty (Gamble 2008; Patton 2010; Westley et al. 2006). As he notes, the contexts of social

---

[16] Patton made this point repeatedly in the Du-Pont/McConnell sponsored workshops on developmental evaluation workshops held in 2005-2006.

innovation and high complexity are not good places for "control freaks" (Zimmerman et al. 2006).

4.8 Unique Challenges

All evaluations typically deal with a variety of practical constraints (e.g. limited time, resources and data), political dynamics and engendering the use of findings. Developmental evaluators much also must confront a number of additional challenges unique in developmental evaluation: power, credibility, a large volume of data, adaptive budgeting and planning, and maintaining a balance between a focus on developmental processes and results.

4.8.1 Power

Power is a central dimension in any emergent situation. Intervention stakeholders have different levels of power – to shape the purpose and design of an intervention as well as the subsequent collection and interpretation of emerging data (Gamble 2008). Patton and his colleagues argue that funders in the philanthropic sector have an extra-ordinary influence on the decisions on the strategy and decisions of grass roots social innovators they fund (Westley et al. 2006). Developmental evaluators need to be sensitive if and when power manifests itself during the development of an intervention and to surface them if it begins to interfere with the use of data and critical thinking in the evaluation process.

4.8.2 Credibility

In order to be part of an emerging process, the developmental evaluator's role is to operate as a team member and be as close to the emerging innovation and innovators as possible so that that they are able to stay in touch with an often unpredictable unfolding of an intervention and be present for "developmental moments", while at the same time being sufficiently external to the process that they can focus on facilitating a disciplined process of data-based decision-making and rigorous thinking (Patton 1999; Gamble 2008).

This is a difficult balancing act. Several participants in Du Pont-McConnell Foundation training in developmental evaluation, for example, reported that in the midst of emergent

situations, they assumed responsibility for taking meeting notes and chairing meetings and even stating preferences for program design.[17]   Even where a developmental evaluator maintains this balance, some people familiar with more traditional approaches to evaluation may feel that s/he is too embedded in the process and too close to the innovators to be objective. As a result, they may find their role, activities and contributions as less credible than they might be.

4.8.3 Volume of Data

Developmental evaluation activities can generate a great volume of data (Gamble 2008). Change agents cast about to get a better handle on the environment. There are lots of different conversations about the issue and how to address it. It includes a lot of false starts and midcourse corrections. The effects of interventions that do unfold are often unpredictable and the net is cast wide. All of these activities generate a pool of data upon which to draw to inform the thinking and decisions of evaluation users. Evaluators and their users often struggle to determine which areas of activities to track and how best to efficiently synthesize and effectively communicate the data that emerge (Patton 2010).

4.8.4 Adaptive Planning and Budgeting

The unpredictable nature of developing interventions makes it difficult for evaluators to design, plan and budget for developmental evaluation. The following account by Al Etmanski of the journey of his group in creating a new approach to supporting persons with disabilities to lead an independent life after their family caregivers had passed away illustrates the messy dynamic of social innovation:

> We had to fight for these concepts. They didn't just slip into your hand and you'd say, "Oh, I think I'll follow that one for a while." We would go over an issue five meetings in a row, agree on something, and then at the sixth meeting decide to go in a completely different direction. We were civil, but there was tension there to figure out what the values meant… There was a lot of comfort in ambiguity… I had just come from a job in which I made twenty decisions a day — probably all of them were bad decisions, but I was able to make them — to a job in which I was expected to make no decisions for months. We persevered, walked around the issue, had a look at it, said, "Okay, this is it; we might as well go in this

---

[17] This is a personal observation based on my involvement in those sessions as a participant.

direction." And then we'd change our minds. And we still do that. We were a learning group. Everyone was curious. (Westley et al. 2006: p. 75.)

The ability to manage the practical challenge of planning and continually adapting an evaluation design that is both effective and feasible is important for the evaluation user as well as the evaluator concerned with working within a budget.[18]

4.8.5 Keeping a Results-Focused

One of the most significant challenges for a developmental evaluator is to avoid focusing solely on the *process* of development rather than the *substance* of what is being developed (Gamble 2008). This is particularly difficult given the reality that social innovators typically expend a great deal of effort trying to create a concrete intervention able to yield measureable outcomes with most of their efforts leading to "dead ends" (Westley et al. 2006). Part of the task of developmental evaluator, therefore, is to periodically point out when nothing is being developed and/or what is being developed is generating very weak effects.[19]

4.9 Developmental Evaluation in the Literature

4.9.1 Other References to Developmental Evaluation

The phrase developmental evaluation is new in the evaluation literature. However, evaluation theorists and practitioners have addressed characteristics of the various niches Patton proposes for developmental evaluation in varying degrees for quite some time. Is the substance of developmental evaluation something authentically new to the field, or is it simply a "repackaging" of existing evaluation thinking and work?

The niche which focuses on the role evaluation can play in conceiving and designing an intervention has been extensively covered in the literature for decades (e.g. Chen 1996; Owen 1999; Todd & Wolpin 2006; United States Government Accountability Office 1990). Patton simply names this pre-formative evaluation and confirms that its purpose is to develop tangible

---

[18] This concern was raised by several audience members of a presentation on developmental evaluation that Michael Patton provided at the 2009 Canadian Evaluation Society meeting in Ottawa.

[19] Patton made this point while reflecting on a case study where the members of a cross-Canadian network devoted to improving social innovation in Canada had failed to create some joint projects after several years of effort.

interventions that eventually will evolve into a formative situation requiring formative evaluation.

There does not appear to be a clear and discrete reference in the evaluation literature to the niche of "ongoing development" of evaluation models in dynamic contexts. In fact, some evaluation practitioners tend to consider ongoing development as a type of formative evaluation with no fixed end point because the focus is on strengthening the intervention.[20] Patton argues that the distinction is real and important: formative evaluation focuses on improving an existing model with the hope of stabilizing it for an evaluation summative evaluation, while developmental evaluation is used when stakeholders are interested in continually reconceptualising and redesigning an intervention in order to reflect changes in their contexts.

The literature on the role of evaluation in the process of scaling up and replicating interventions tends to reflect two broad orientations (van Oudenhoven & Rekha 1998). The universalist approach reflects a strong positivist bent and is rooted in the idea that if and when an intervention has been stabilized and tested through a rigorous summative evaluation, evaluators can assist stakeholders to codify its key features and then switch to a form of "fidelity evaluation" in which they work with would-be replicators to implement the model with fidelity to the codified design. Advocates of the contextualist approach, on the other hand, argue that social interventions and their outcomes are highly shaped by the social, economic and political context in which they emerge. As result, they prescribe against the simplistic urge to replicate the exact model of an intervention to other contexts and instead (a) help intervention stakeholders surface and distil "transferable lessons" from the intervention and then (b) assist would-be replicators to adapt these lessons to the realities of other unique contexts (Pawson & Tilley 1997; Schorr 1997; van Oudenhoven & Wair 1998). While his pragmatist approach suggests that he would choose the approaches more suitable to a given situation, Patton's description of the role of developmental evaluation appears to be more contextualist than universalist.

---

[20] This reflects the author's observation of two discussions amongst practitioners: participants of the Du Pont McConnell Family Foundation training sessions on Developmental Evaluation as well as the questions raised by participants of a workshop on Developmental Evaluation by Patton at the Canadian Evaluation Society in Ottawa in 2009.

Evaluation theorists and practitioners have been working on assisting policy makers and program designers to assess diverse efforts at systems changes, such as neighbourhood renewal, welfare reform, economic development, ever since the evaluation field began to grow dramatically in the 1950s. Jane Jacobs was one of the earliest advocates of a new approach to evaluation when she criticized urban planners for employing faulty science and evaluation in their assessment of neighbourhood renewal efforts (Jacobs 1961). Since then, evaluation theorists have explored and developed new evaluative approaches that deal with issues related to complexity (Williams & Imam 2007), comprehensive approaches (Kubsich, et al. 2002) and multi-site cross scale initiatives (The Aspen Institute 1995, 1997a). Patton acknowledges and summarizes the variety of work in this area without appearing to add anything substantively new (Patton 2010).

Finally, the role of evaluation in crisis situations is covered more by observers and theorists in the field of organizational development and management, than in evaluation circles. Building on the earlier work of Simon (1956), recent books such as **Blink** (Gladwell 2005), **Sway** (Brofman & Brofman 2009), **Nudge** (Thaler & Sunstein 2009) and **Predictably Irrational** (Ariely 2009) explore how cognitive biases and data shape the thinking and decisions of organizations and leaders in crisis and high stakes situations. Evaluators would describe many of the prescriptions offered to avoid bad decisions by the authors of **Think Again** (e.g. assisting decision-makers gather and interpret data in different ways) as evaluation activities (Finkelstein, Whitehead & Campbell 2008). Patton, however, appears to be the only major evaluation theorist to argue that these situations represents a special niche for evaluators and that evaluation practices should be customized to match the unique characteristics of these contexts.

In summary, Patton seems to do more than pour old evaluation wine into new bottles. First, he argues that for each niche the primary purpose of evaluation is to assist intervention stakeholders to "develop" – rather than improve or judge -- the intervention using data and critical thinking, a distinction that does not appear to be made by any other evaluator theorist or practitioner. Next, Patton spends a great deal of time describing the messy and unpredictable dynamics of the development process in order to sensitize evaluators to the need to be adaptable and flexible in their approach. Finally, Patton's utilization-focused and contingency approach to methods challenges evaluators to resist the urge to develop a standard set of tools and approaches

(e.g. the universalist approach to replication which highly favours the use of randomized controlled trials regardless of context) but rather tailor them to the situation and preferences and the idiosyncratic needs and preferences of evaluation users.

4.9.2 Comparing Developmental and Traditional Evaluation

While Patton argues that the critical distinction between developmental evaluation and the more traditional forms of formative and summative evaluation is the distinction between purposes: i.e. development, improvement and judging or merit or worth, he identifies nine major other philosophical and practical distinctions as well (See 4-3).[21]

4.9.3 Critiques of Developmental Evaluation

There is no published written critique about developmental evaluation in a major or known publication as of June 2010, when the literature review of developmental evaluation was completed. The author was unable to find any direct or indirect reference to neither developmental evaluation in either the Canadian Journal of Program Evaluation nor the American Journal of Program Evaluation. Michael Quinn Patton reports that, to his surprise, he is not aware of any published critique either.

Patton does share critiques he has come across in his own work. The first is that developmental evaluation may not be evaluation at all, but simply another form of organizational development (Patton 1999)[22] In his early writing on the topic, he is sympathetic to this perspective:

---

[21] In his book on developmental evaluation, Patton expands the number of distinctions to twenty three. I did not have this list of distinctions until the interviews were underway with study participants.
[22] Several of the participants in Developmental Evaluation Community of Practice established by the Du-Pont Canada the J.W. McConnell Foundation described developmental evaluator role as that of a coach. Personal Observation as member of the Community of Practice.

Table 4.2: Comparing Traditional and Development Evaluation

TRADITIONAL EVALUATIONS	COMPLEXITY-BASED, DEVELOPMENTAL EVALUATIONS
Render definitive judgments of success or failure.	Provide feedback, generate learnings, support direction or affirm changes in direction.
Measure success against pre-determined goals.	Develop new measures and monitoring mechanisms as goals emerge and evolve.
Position the evaluator outside to assure independence and objectivity.	Position evaluation as an internal, team function integrated into action and ongoing interpretive processes.
Design the evaluation based on linear cause-effect logic models.	Design the evaluation to capture system dynamics, interdependencies, and emergent interconnections.
Aim to produce generalizable findings across time and space.	Aim to produce context-specific understandings that inform ongoing innovation.
Accountability focused on and directed to eternal authorities and funders.	Accountability centered on the innovators' deep sense of fundamental values and commitments.
Accountability to control and locate blame for failures.	Learning to respond to lack of control and stay in touch with what's unfolding and thereby respond to strategically.
Evaluator controls the evaluation and determines the design based on the evaluator's perspective on what is important.	Evaluator collaborates in the change effort to design a process that matches philosophically and organizationally
Evaluation engenders fear of failure.	Evaluation supports hunger for learning.

(Source: Patton 2006, p. 30)

I won't quarrel with that. There are sound arguments for defining evaluation narrowly in order to distinguish genuinely evaluative efforts form other kinds or of organizational mucking around. But, in each of the examples I have shared, and there are many others, my participation, identity and role were considered evaluative by those with whom I was engaged (and by whom I was paid). […]What we lose in conceptual clarity and purity with regard to a narrow definition of evaluation that focuses only on judging merit or worth, we gain in appreciation for evaluation expertise (Patton 1999, pp.111-112).

Over time, Patton has become more emphatic in arguing that developmental evaluation is a legitimate type of evaluation. In his latest edition of ***Utilization-focused Evaluation*** (2008), he argues that the orientation and practices normally employed in summative and formative evaluation are inappropriate in developmental situations:

None of these traditional criteria are appropriate or even meaningful for highly volatile environments, systems-change-oriented interventions, and emergent social innovations (Patton 2008, pp.137).

Patton goes on to describe developmental evaluation as one of six major evaluation purposes to which evaluative thinking, principles and methods, and expertise can employed alongside formative, summative, knowledge generating, monitoring and accountability-focused (Patton 2008, p. 139).

A second major critique is that the distinction between 'formative' and developmental' situations, that is, the line between developing and improving an intervention, is simply a matter of degree and therefore developmental evaluation is possibly a subset of formative evaluation.[23] Patton argues that there is a qualitative distinction between formative and developmental situations and describes five indicators when a developmental situation is entering the "formative zone". These are:

- There is a sense that it is time and timely to move from divergence (generating options) to convergence (placing best bets).

- There is a sufficiently well formulated intervention that can be conceptualized, implemented, tested and improved.

- One can also conceptualize what the ultimate summative questions and issues would be to inform the formative evaluation.

- Key stakeholders are energized by the intervention and want to take it to the 'next level'.

- There is funding for quality formative implementation and evaluation, and a commitment to carry out a summative evaluation.[24]

Both these critiques were provided in response to Patton's early writings and presentations of developmental evaluation. The volume, diversity and substance of critical feedback may increase with the publication of a full book on the topic.

---

[23] Personal conversation between author and Patton. January 19, 2010.
[24] Personal conversation between author and Patton on January 28, 2010.

4.10 Documentation of Practice

The members of the evaluation community and users of evaluation know a lot more about developmental evaluation and its characteristics today than they did sixteen years ago when Patton published his first article on the topic. Patton estimates that since the mid-1990s, he has given approximately forty speeches, fifty workshops and seventy-five general speeches on development evaluation.[25]

Despite this broad dissemination of information on the approach, evaluators and evaluator users know very little about the practice of developmental evaluation. Gamble's (2008) ***Developmental Evaluation*** primer is based on the initial experiences of a small group of people who attended a workshop series on developmental evaluation delivered by Patton in 2005-06. Beyond this, there is no informal or formal research into the use of developmental evaluation beyond that undertaken and communicated by Patton himself.

In the spirit of contributing to the empirical knowledge on evaluation practices, this study sets out to research and document the practices of "early adopters" of developmental evaluation.

---

[25] Personal conversation between author and Patton on January 17, 2010.

Chapter Five: Results

5.0 Introduction

The purpose of this chapter is to describe study participants' experiences and reflections on developmental evaluation. When and where appropriate, these results are compared with the emerging theory of developmental evaluation.

5.1. Purposes

Patton's central argument for developmental evaluation is that it distinct from more traditional forms of evaluation and others non-evaluation disciplines such as organizational development. Most study participants appear to agree with these broad distinctions. Some participants, however, are unclear about the distinction between developmental evaluation and other disciplines.

5.1.1 Evaluation Distinctions

Patton argues that the purpose of *formative* evaluation to improve, refine and stabilize a model or intervention; the purpose of *summative* evaluation is to judge the merit or worth of a fixed model or intervention in order to determine whether it should be discontinued, sustained and/or replicated; and, the purpose of *developmental* evaluation is facilitate the development of a model or intervention, by creating an entirely new one and/or continually adapting/restructuring an existing one.

Study participants identified a range of features they felt characterized developmental evaluation and distinguished it from other types of evaluations and disciplines (See Table 5-1).

Table 5-1 Developmental Evaluation Distinctions

KEY FEATURES	PARTICIPANTS	N=18
Developmental Evaluation		
Purpose	A, E, F, H, I, J, K, L, M, N, P, Q	66%
Specific Niche	K, D	11%
Relationship with Users	B, M	11%
Summative-Developmental Distinction		
Purpose	A, B, D, E, F, G, H, I, J, K, L, P, Q, R	78%
Methods	A, D	17%
Relationship with Users	F, N, P	11%
Formative-Developmental Distinction		
Purpose	D, E, K, M, Q, R	33%
Clarity of the Intervention	A, B, I, L O, Q	33%
User Expectation of Evaluand	A, B, M, Q	22%
Approach to Learning	F, G, C	17%
Other Distinctions		
Other Methods of Inquiry	C, B, L, O	22%

(a) developmental evaluation

The majority of study participants (A, I, E, F, H, J, K, L, M, N, P, Q) reported that they felt that the purpose of developmental evaluation was to support the "development" of an intervention. This includes:

- a conceptualization of the problem (A, C, G),

- agreement on of desired outcomes (A, E)

- choice/creation of strategy, tactics, design, (A, E, G, J), and

- practical elements of an intervention or way of operating (A, G, R).

While many participants identified one or more of these elements of an emerging intervention as "developments", several stressed that the developmental evaluation should ideally help produce an overall workable intervention. Study participant I, for example, described the outcome of a developmental process as "something both conceptual and practical,

like a program, a policy or a response to a crisis". Study participant F further emphasized the development of something "actionable":

> Developmental evaluation helps to support the development of an actionable focus and direction and helps groups test and refine their understanding of both the initiative and the larger landscape in which it is situated.

Several study participants focused on the type situation in which evaluative thinking and activities were employed. Study participant K remarked that developmental evaluation was "a way to work evaluatively within adaptive, dynamic situations". Study participant D reported that the aim of developmental evaluation was "to be helpful in situations of complexity and innovation [through] the deliberate gathering of evidence".

Two participants pointed to the unique relationship between the evaluator and the evaluation users. Study participant B noted that developmental evaluation is "a long term partnering relationships between the evaluator and those engaged in innovative initiatives and development" while study participant M emphasized that the evaluator was "part of the team".

(b) summative-developmental evaluation

All study participants reported that they felt there was a clear distinction between summative and development evaluation. The majority of participants (A, B, D, E, F, G, H, I, J, K, L, P, Q, R) reported that the greatest difference related to the purpose of evaluation: summative evaluation is about judging the merit or worth of an intervention while the purpose of developmental evaluation is to create or restructure an area of work. Study participant H noted:

> Summative evaluation is about exploring whether we are achieving the outcomes we say, determining whether our approach works, deciding whether we should scale it up. Developmental is focusing on surfacing what we are trying to achieve, what we are learning, the dynamics of systems change, being creative and innovative.

Study participants M, N and O remarked that the relationship between the evaluator and initiative stakeholders is also unique. In developmental evaluation, the evaluator is considered part of the team developing or restructuring the intervention while in summative evaluation they are external to the team.

Two participants (A, D) felt that summative evaluation placed a much stronger emphasis on employing quantitative methods to capture the effects of interventions while developmental evaluation placed a broader emphasis on mixed or qualitative methods.

(c) formative-developmental evaluation

Study participants varied in their clarity about the nature or degree of the distinction between formative and developmental evaluation. While they identified three broad areas in which the two approaches to evaluation were different, several participants also expressed confusion or disagreement about whether the differences were clear or meaningful.

Several participants A, B, I, L, O, Q reported that they felt developmental and formative situations differed in the degree to which the intervention being evaluated is conceptualized and/or had a developed a "mechanism for change". Study participant B reported:

> [Groups] almost always they have a pretty inadequate vision and focus for what they are doing so to that extent most are what you would call formative or developmental, depending on how far along the conceptualization is.

In developmental situations, study participants felt that evaluation users typically are working with only a general sense of the challenge they are trying to address, the broad strokes of what they wanted to accomplish, and a "hunch" (Study Participant O) of how to get there. In formative evaluation, the evaluation users are working with a fairly well articulated intervention, often created through an earlier round of experimentation. Other study participants framed the distinction in slightly different ways:

> Formative evaluation is about improving or enhancing the quality of what you are already doing. (Study Participant A)

> Formative evaluation is when most of the logic is nailed down but we need to improve it and its practice. Developmental evaluation is to inform the creation of the logic and practice. (Study Participant I)

> When we talk about formative, it is usually to put into place a structure, process and system that keeps you roughly aligned with your intervention. (Study Participant Q)

Many study participants (D, E, K, M, Q, R) reported that developmental and formative evaluation had distinct *end points*. Whereas a strict formative evaluation was designed to help improve and eventually stabilize a model for a summative evaluation, developmental evaluation was employed when the intervention stakeholders were open to continual – even radical – changes to an intervention. This is illustrated in the following responses:

> If a group is really trying to define a model that is at least stationary enough that it can be (a) defined and (b) evaluated as model, then is formative. When you are with an organization that is not that clear on what they what they want to achieve and/or what their model is, that is developmental. (Study Participant M)

> Formative evaluation is about modifying things to eventually get a fixed design. Developmental evaluation is the opportunity for open-ended evolution and development of what is being evaluated. (Study Participant D)

Finally, several study participants described how formative and developmental evaluation emphasized different levels or types of learning processes. Study participant F argued formative evaluation typically focused on "conventional" learning while developmental evaluation tended to involve "adaptive learning". Using an analogy to illustrate the distinction, participant F stated:

> Most of the world functions on the basis of conventional learning which is quick and easy – you just follow the patterns that have already been set. Adaptive learning is a whole other ball game. It is slow and effortful. I sometimes use the analogy of a DVD to illustrate the difference between the two kinds of learning. Everyone knows how to use their DVD – some of us can even program it. But what if it breaks, or what if we want to design a better model? The kind of learning and understanding that is required for diagnostics and design is very different from the kind of learning that is required to use something. Most of what we do is at the conventional level and we don't have a lot of tolerance for the kind of learning and development that is required to support diagnostics and design.

In a similar vein, study participant G explored the difference between the different types of thinking and behavior typically employed in development and formative situations:

> I do think that there is a substantive difference in the dynamic of creating – or restructuring -- an enterprise and the process of refining and improving an existing one. The process of development is largely inductive or abductive, where you are weaving together and experimenting with fragments of ideas and experiences in a real world setting hoping that they eventually come together in some type of coherent way of looking at the issue and changing something. Formative evaluation is essentially a deductive process, where the task is make

an existing approach better, but not essentially different. Development work goes deeper and is messier and more unpredictable than formative work.

Study participants I also felt that there was a distinct difference between formative and developmental evaluation but struggled to describe it precisely:

> I believe there is a difference, but I can't really put my finger on it. I think it has something to do about the differences between the act of creation and the act of improving. In formative situations, you are already working with a center of gravity, a way of thinking and a rough pattern of activities and the task is to refine or improve that pattern. This usually leads only to incremental changes. In developmental situations, you don't yet have much of anything – you need to create a perspective(s) and pattern based on hunches and ideas. Or, you may be working with an existing pattern than may want or need to radically restructure meaning you have to go back to the drawing board and perhaps start from zero. More is on the table in developmental evaluation than in formative evaluation, you search more broadly and more deeply, there is a lot more exploration, probing and testing, all of which could lead to entirely new, even radical, approaches. It seems to me that these are qualitatively different situations and have different implications for innovators and evaluators.

While the majority of participants described three broad differences between formative and developmental evaluation, not all study participants are sure that the distinction is clear and/or sufficiently meaningful to distinguish the two. Study participant C argued:

> I think that the developmental and formative distinction is somewhat arbitrary. Maybe I have not read the right books! I was surprised to hear that distinction. If DE is about supporting emergent development and formative improvement oriented, the reality is that there are very few truly formative situations. I actually like the distinction – I know why it might be important - but it's unfortunate to degrade formative.

Study participants B and Q shared a similar sentiment. Study participant Q stated, "I don't use these terms because they don't mean much to people outside the area". Study B went further: "In general, I still don't think in the categories of developmental versus formative. It's not a major issue for me. Instead, I look for where the group is at and how I can help." While three participants were not convinced that the distinction between the two was substantive enough to warrant a different label, two reported that Patton's elaboration of the process of developmental evaluation was a useful contribution to the field.

5.1.2 Emergent Questions & Issues

While all the study participants acknowledged that developmental evaluation was a new concept and term, many of them reported that they felt that they had employed a version of developmental evaluation in their work for some time.

(a) relation to other inquiry methodologies

Patton is clear to point out that developmental evaluation has generally focused on how it differs from formative and summative evaluation. Several participants raised questions and/or expressed confusion about other evaluation approaches and/or particular methodologies.

Some study participants felt that the practice of developmental evaluation was very similar to the methodological approaches of action research (Participant C), reflective practice (Participants B, O) and appreciative inquiry (Participant O).

Study participants L and M declared that they either saw similarities with between developmental evaluation and the practice of evaluation monitoring employed in the field of international development and wondered if there were any substantial differences between the two activities.

(b) relation to planning

Study participants F, H and K described developmental evaluation as something they used as part of their work in urban and regional planning. Study participant K, a full time evaluator with formal training in urban and regional planning, argued that planning and developmental evaluation are overlapping disciplines:

> I have been involved in systems focused evaluation before – much of it from a planning background. There is not a lot of difference in my opinion. I have always done this – as a planner – through an evaluative door. DE is a dialogue between planning and evaluation. People like to have people with them helping to do their work better. There has always been a hunger for that work: i.e. evaluative planning.

Study participant F and H felt that developmental evaluation was especially appropriate when planning to address complex – as opposed to simple – issues where linear approaches of problem solving were not as appropriate as iterative and adaptive ones.

(c) relation to organizational development

Study participants D H, N, O remarked that developmental evaluation and organizational development appeared to be overlapping disciplines, where the evaluator and organizational development facilitator operated as part of the team and fulfilled many of same functions, such as facilitation, conflict resolution and strategy development.

Study participant D, with many years of facilitating teams in a large corporation develop or refine strategies or products, described how s/he fulfilled a quasi-evaluation function in her work without calling it evaluation:

> I would not have considered myself an evaluator – or completed formal evaluation work – at all prior to [participating in the DE training]. At least, I would not have called it evaluation. I did, however, do a lot of facilitation with teams within our company coming up with new strategies. As part of that, I was used to asking question – often tough questions. […] We found the idea of more deliberately gathering and considering evidence – hard and soft – as something we had not emphasized as strongly in our reflective practice in the past.

Study participants A, E, F and K observed that developmental evaluation was a natural evolution of "adaptive leadership", "adaptive learning" and/or "adaptive management", an overall orientation to planning and decision-making in complex issues, uncertain and ever-changing environments, where the emphasis is monitoring the environment, "learn-by-doing", iterative decision-making and flexible responses.

Study participant E, a professional evaluator who specializes in assisting coalitions develop and evaluate advocacy campaigns to change public policies, felt that development evaluation was very similar to the broader concept of "strategic learning", a term developed and

popularized by The Atlantic Philanthropies that emphasizes building up the capacity of grant recipients to assess progress and learn from their work.[26]

Study participant K and M described developmental evaluation, with its emphasis on continual development, shared many of the same characteristics of a "learning organization" which strives to involve members of an organization to be very deliberate in learning about and adapting their work to better fulfill their mission or achieve a particular outcome.

The majority of study participants who described the overlap between developmental evaluation and organizational development reported that developmental evaluation was distinct in its emphasis on the "rigorous use of data" and/or "critical thinking" in decision-making.

(d) relation to developmental practice

While study participants were relatively clear about the similarities and differences between other disciplines and developmental evaluation, study participant A raised a question about the possible overlap with developmental practice, a particular approach to managing projects and programs in developing countries.[27] She notes: "There is a lot of similarities between DE and developmental practice. I am not sure where one begins and the other end".

5.2 Niches

Patton has most recently has come to describe five situations in which he feels it may be suitable to employ developmental evaluation. These include:

- Adapting an intervention to new conditions in complex dynamic systems.

- Adapting models or general principles to a new context as ideas and innovations are taken from elsewhere and developed within a new setting.

- Developing a rapid response in the face of a sudden major change or a crisis (e.g. , a natural disaster or financial meltdown).

---

[26] See weblink: http://www.atlanticphilanthropies.org/strategic-learning

[27] Developmental practice is a term popularized by The Community Development Resource Centre (aka, the Centre for Developmental Practice), based in South Africa, which aims to support organizational innovation.

▪ Developing a new and emerging innovation.

He also acknowledges the likelihood that a sixth, more complex, situation requires the concurrent use of developmental, formative and summative evaluation concurrently, an approach he termed "patch evaluation".

Study participants report that they have deliberately employed developmental evaluation in four of these situations, have no experience in two of them, and have surfaced a seventh possible niche: an evaluative probe (See Table 5.2)[28]

Table 5-2 Niches for Developmental Evaluation

Current Niches Envisioned For Developmental Evaluation
{
Unrealized Niche
* Rapid Response
* Replication

Intended Niche
* Pre-formative
* Ongoing development
* Systems change

Emergent Niche
* Evaluation probe
}
Study Participants Examples
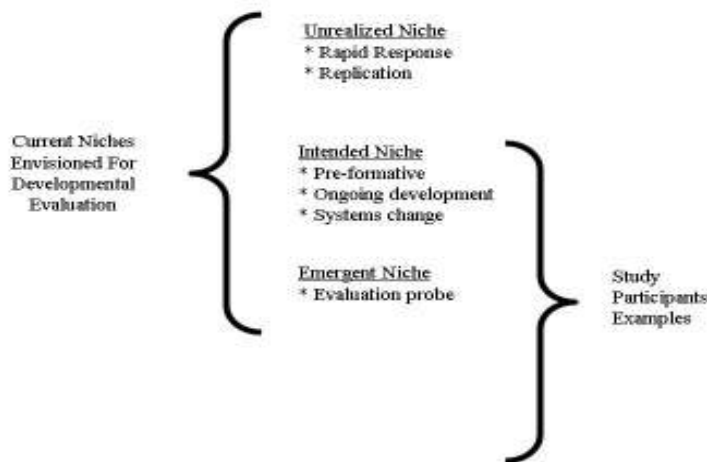
5.2.1 Realized Niches

(a) pre-formative

In his writing on social innovation, Patton (2007) explored how developmental evaluation could be usefully employed to inform the early development of an intervention an iterative process of learning-by-doing and reflection, which would eventually lead to broadly

---

[28] The idea to use this simple concept to visually present the research results in this area originally emerged during the review of the literature on planning and evaluation and again when reading a manuscript of Patton's book on developmental evaluation.

conceptualized theory and mechanism for change. Once intervention stakeholders had settled on broad features of a policy or program, they would employ a formative evaluation to help improve the model. Once the model was fixed, it would enter into a multi-year summative evaluation where the emphasis was rigorously tracking the effects of the intervention so that it could eventually be subjected to a summative evaluation, usually managed by an external evaluator, in order to judge its merit or worth and determine whether it should continued and/or scaled up.

Many study participants reported that they were involved in many pre-formative situations (A, B, C, D, E, F, H, I, N). Study participant D, for example, described being involved in one pre-formative situation while study participant B reported being involved in up to 50 developmental evaluations, many of which were pre-formative.

Study Participant D's experience assisting the staff in a science museum develop and test a way of putting together exhibits that encouraged innovative thinking and behaviour amongst museum attendees illustrates (a) developmental evaluation in a pre-formative situation as well as (b) a rapid transition to formative and summative assessment. In the early days of the program, the study participant worked with staff to create a list of attributes they constituted characteristics of innovative behaviour, which they called the 'innovation framework'. They then used a process of rapid prototyping to develop and evaluate new exhibits based on data that they gathered and analyzed daily. This led to immediate – sometimes radical -- changes in the museum exhibits as well as the group's innovation framework.

After a very intense start-up period, the number and scale of changes in exhibits and innovation and framework decreased and a relatively stable pattern of exhibits and conception of innovation emerged. At this point, the group began to refine – rather than change – their program. Study participant D reports:

> Around, 2007-08, the first round of exhibits began to move from prototypes to relatively stable exhibits. The evaluation then focused more on helping the group improve or refine the exhibit than developing it.

The formative phase did not last long. S/he went on to describe how even though staff continued to make changes to the exhibit, the managers at the museum declared that they wanted to know the whether the program "worked" and therefore warranted continued

86

investment. As a result, s/he and the staff team turned their efforts to tracking and analyzing the effects of the exhibits on museum attendees. While the emphasis was on making a summative assessment of the program, study participant D did not regard it as a proper summative evaluation:

> This was a summative moment – rather than a summative evaluation – given that that the innovativeness framework did not stabilize sufficiently for a summative evaluation. Nor was it an exhaustive assessment - we realized that we'd like to go back and gather more data but were unable to due to the lack of resources.

While the end-of-project results were sufficiently encouraging that the museum management continues to use and adapt the innovation framework to this day, however, the series of exhibits that drove the development of the framework were discontinued due to a combination of staff turnover and a shortage of resources.

The other examples using developmental evaluation in pre-formative situations include: study participant F described her/his work with a collaboration or agencies develop a new program to assist at-risk kids improve their attendance at school; study participant N assisted the leadership of faith based organization create a new model of community service that could be integrated into the creation of a new facility in an urban neighbourhood; study participant A assumed the role of developmental evaluator in an international project to assist local farmers develop a new techniques for dry-land agriculture in the Caribbean.

(b) ongoing development

Several study participants also provided examples of using developmental evaluation to inform the ongoing development and adaptation of a policy or program operating in a dynamic context. Study participant D described developmental evaluation in this way:

> If formative is developing a program then stabilizing it and summative evaluation is about implementing with a high degree of fidelity, then developmental evaluative is taking the formative process with no end point.

Table 5-3
Developmental-Formative-Summative Evaluation Progression

The examples provided by study participants in this research suggest that a progression of developmental, formative and summative evaluation may unfold ideally only in special conditions.

There were few examples of full progressions. Study participants B and K observed that because of their cost and required expertise, the strict pre-formative, formative and summative progression tend play out for large scale, well-funded projects that typically employ universities and or research organizations as the evaluators.

Study participants B, D and K describe examples of being involved in developmental-formative-summative evaluations where the model was not-yet-stable or only briefly stable before the emphasis shifted towards summative judgements and where they same evaluator was involved in each phase (as opposed to having an external evaluator carry out the summative evaluation).

Some types of interventions are apt to require continual adaptation. Study participant E, for example, felt that the majority of her/his work to assist advocacy coalitions evaluate their efforts to change public policy was unavoidably open-ended, particularly the longer the time frame in which the group worked:

> Most of my work is related to advocacy work, projects and initiatives. These tend to have what you might call characteristics of development situations: they involve multiple players with different perspectives and organizational challenges, operate at the levels both of broad goals and directions and very specific strategies, and sometimes have time frames of up to fifteen years.

S/he went on to a describe how the dynamics of developing, testing and managing advocacy work required an ebb and flow of both developmental and formative evaluation: after a great deal of effort, a group may settle on a particular goal and strategy and begin to testing it on the ground, only to have to revisit it – and possibly rethink it entirely -- when their initial efforts proved (un) successful and/or the strategy no longer seemed to fit the realities of the fast moving environment in which they were operating.

Other interventions may require find a need to adapt their work only periodically. Study participant O, for example, was involved in assisting a faith-based organization explore how it might innovative with its long-established model of supporting persons with disabilities. Study participant H assumed the role of developmental evaluator for an international human rights

organization that became interested in exploring dramatically new ways to engage volunteers in the work of the organization.

(c) cross scale systems change

Study participants also provided examples of their efforts to provide evaluation support to organizations or groups seeking to change the systems underlying complex social, economic and environmental issues and require assistance in developing a coherent and plausible conceptualization of their work, rapid feedback on effects of their efforts, and critical thinking in adapting their strategy to reflect new learning's and context.

Study participants I, for example, provided evaluation support to a network of grant making organizations interested in supporting the rapidly expanding movement of community-based resource management and conservation in a coastal region of North America. The role of the evaluator was to provide feedback on the extent to which (a) the network members were operating according to their principles and (b) to surface the factors that enabled successful collaboration with the growing number of local partners so that these could be integrated more clearly in the strategy to expand to other communities.

Study participant M worked with a large number of local health collaborations who were focused on improving the capacity of rural health systems to provide quality home visiting services. The role of her/his organization is to help local coalitions develop a theory of change of their approach, track the effects of their effort, and monitor how continual changes in the environment (e.g. Federal policy) affect their work.

Study participant G provided periodic evaluation support to a leadership group of private sector, public, non-profit and community leaders on their efforts to reduce the community's poverty rate by fifty percent within ten years by making changes in a variety of 'systems': e.g. education, housing, early childhood development, welfare policy, and workforce development. His/her role was similar to one played by study participant M in changing rural health systems, and focused on helping the group develop an initial a theory of change, to continually capture and analyze the broad effects of their outcomes, and periodically reviewing whether changes in

the environment, the arrival new actors or the emergence of new learning's prompted the group to think about their work differently or elect to adopt new strategies.

All of these programs happened to be cross scale in nature that involved multiple sites at the local level, regional, provincial/state, national and international level.

(d) patch evaluation

Several participants shared examples where they employed developmental, formative and summative oriented work concurrently. When study participant M described the six evaluation projects s/he was working on the week of the interview, s/he determined that of five of them included some element of developmental evaluation, and three involved a mix of developmental, formative and summative evaluation:

> Most of my work is patch evaluation. I think [Michael Patton] called it briceloge. I don't force – or even believe in - a pure developmental, formative or summative evaluation. I certainly don't enough power or credibility to make that happen!

Study participant M continued to describe her participation in a federally funded health care project operating in seven diverse rural regions that was integrated elements of a summative and developmental evaluation. This included an experimental design methodology to assess the effects of a home visiting program on client's well-being. At the same time, s/he was also working in each community to assist each local health coalitions' organic process of shaping and evaluating strategies to re-shape local policies and systems affecting home visiting. Each of the local processes was organic, adapting to new learning's and responding to a time of turbulence in federal health policies, as well as unique, reflecting the unique political and social contexts in which each group was operating

If the scope of the initiative and evaluation is large enough, patch evaluations may involve more than one evaluator/evaluation group working together or independently. In the following example, study participant I describes her/his discussions with a large non-profit organization involved in multiple evaluation processes.

I was invited in and spend some time talking about DE with [a large non-profit organization providing services to homeless persons] They were excited to do [DE] without really grasping what it was and were concurrently doing other evaluative work: e.g. [a] University was doing social impact assessment, another group was helping them re-jig their organizational structure to strengthen cross fertilization of learning.

Other interviewee shared examples of patch evaluation. Study participant F described providing developmental evaluation support to a collaborative group working to keep at-risk youth in school while another evaluator was concurrently working on developing a social return on investment analysis to create a baseline of the economic and social costs of the intervention for use in an eventual summative evaluation. Study participant R described her role in evaluating an expanding network of youth leadership programs in an organization alongside three other evaluative processes.

5.2.2 Unrealized Niches

None of the eighteen participants in this study had experience using developmental evaluation in two of Patton's five niches.

There was only one interviewee who identified working in a crisis situation, where decision-makers need to employ evaluative thinking in fast chaotic situations, as a niche for development evaluation. When encouraged to describe this situation in more depth, however, s/he noted that s/he had never deliberately employed developmental evaluation in a crisis situation, but described providing the management of aid agencies providing emergency local assistance to people in Haiti after the aftermath of the 2010 earthquake was a good illustration of where developmental evaluation might be useful.

None of the persons interviewed for this study identified the replication of principles of proven models from one context as a niche for developmental evaluation, nor did they share an obvious example of working in this situation.[29]

---

[29] The researcher in this study did approach someone who had earlier reported her experience employing developmental evaluation to help with the replication of interventions from one context to the next, but s/he was declined to participate due to time constraints.

5.2.3 Emerging Niche

One of the unanticipated findings in the interviews was a claim by study participants that developmental evaluation could be used as a "probe" into an existing pattern of activities in the hope that by asking deeper questions about the assumptions behind the work it might encourage program managers to be more creative in their thinking and responses.

Study participant R described an example where a large local foundation contracted her/him to provide evaluation support to several organizations receiving money to manage a multi-year program to support community organizing in a large Canadian city. While the formal purpose of the evaluation effort was to improve the evaluation capacity of the organizations, s/he felt that the real objective was to "shake things up" enough that grantees would consider alternative ways of supporting social mobilization and sensitive issues in inter-cultural integration:

> If I look at my work at [the funding organization], they definitively had two objectives. On the one hand, they wanted [their grantees] to simply do more, better evaluation, so it was about getting better information and using it. But, what they really wanted to do was to turn the initiative on its head: they had invested in this project for 10 years and were not sure that it was working they way it should or could. They were interested in the possibility of completing putting it apart and rebuilding it. They did not say it but it might have been their ultimate goal: i.e. "[the evaluator] can tell us what is really going on we can use that information to make changes". So, they were using the evaluation contract to solve bigger organizational issues.

The subsequent evaluation was "messy" and "awkward" but did eventually sufficiently disrupt the normal pattern of program delivery and funder-grantee interaction that they did begin experimenting with new ways of mobilizing vulnerable residents and adjust the manner in which the funder and grantees worked together.

5.3. Roles & Relationships

Gamble (2008) identifies three broad roles for the evaluator: framing issues, testing quick iterations, and tracking the trajectory of an emerging intervention. The experience of study participants reflects – and elaborates – on these three broad roles and surfaces a debate and

questions about two more roles: judging the quality of the emerging developments and getting involved in matters related to implementation. (See Table 5-4).

5.3.1 Framing Issues

Study participants provided a number of examples of how they helped innovators frame key issues related to the work.

This first was helping intervention stakeholders to define the challenge, problem or issue they are trying to address. Study participant G described how defining poverty was a central task for the evaluation team of a national network of poverty reduction initiatives:

> How you frame a problem matters a great deal. A group that defines poverty as a lack of employment income will focus on strategies that help the poor complete education and get a good paying job. A group that defines poverty as lack of income – without regard to its source - may be inclined to include strategies that expand peoples' access to government income support programs. A group that defines poverty as exclusion from the day-to-day life in the community and access to basic necessities for life, will seek to expand opportunities for civic and political engagement, access to services, food, shelter, clothing etc. And so on.

The study participant went on to describe how the initiative's evaluation team encouraged local groups to develop an explicit and shared "working definition of poverty" to guide their work. They supported this process by facilitating conversations about members' opinions on the characteristics of poverty, creating a summary of typical definitions of poverty and their possible implications for strategy, facilitation evaluation and communication, and by using simulations where group members experienced different dynamics of living in poverty and/or using different definitions in their work. Once a group adopted a definition and began experimenting with strategies and projects, they and the evaluation team would periodically stop to reflect on whether the definition should be changed to reflect new learning, new partners and shifts in the environment in which they operated.

Evaluators can also help innovators frame the outcome to which they would like to contribute. In the case of a science museum where staff wanted their exhibits to encourage innovation amongst their visitors, for instance, study participant D helped the group develop a list

of characteristics of innovative thinking and behaviour based on their own understanding innovation as well as research into the literature on the topic, which they eventually called an "innovation framework". As the staff experimented with new ways to manage their exhibits over the next year, they reflected upon and adapted the framework based on observations of – and feedback from – people attending their exhibits.

Finally, evaluators can also help innovators to better frame and describe the emerging intervention they hope will change the problem they are trying to address. Study participants B and K, both experienced evaluators, reported that this is perhaps the most frequent need they encounter with evaluation users. Study participant B noted that "[*My clients*] *almost always have a pretty inadequate vision and focus for what they are doing",* prompting her/him to work with them to prepare a stronger conceptualization and mechanism for change. Study participant K described a similar experience:

> Ninety-percent of the time people don't have a theory. In fact, I am thrilled and shocked when they do. They are often doing a lot but not conceptualizing the thinking behind the work. They often think they have a theory, but in fact it's really a workplan.

S/he continued by arguing that evaluator was positioned to help program designers improve the "face validity"' of the theory of change underlying their emerging intervention. This included two elements of the theory: (a) the degree to which the assumptions about the cause and effect relationships required to generate an outcome or a desired change were reasonable, and (b) the 'theory of action', that is, the extent to which practical mechanics were likely to realize the theory of change.

5.3.2 Testing Quick Iterations

A second role in which study participants have played a role in developmental situations is to help innovators test quick iterations of their ideas, beliefs and hunches by gathering and interpreting data related to their probes and experiments, drawing conclusions about that data, and using the data to make decisions about further developments in the intervention.

(a) gathering data

Study participants described a variety of ways in which they assisted evaluation users to gather and analyze data on the early effects and implementation of their emergent work.

Sometimes study participants assume full responsibility for the process. Study participant M, for instance, designed and implemented a consultation process that used focus groups to surface the opinions of local residents on types of services a faith based organization might offer in an urban neighborhood. Study participants B and K employed surveys of persons involved in organizations and networks gather their stories of "most significant changes", "problems" or "failures".

In some situations, study participants cooperated with evaluation users to gather and analyze data. For example, study participant P described his role in assisting the organizations involved in an international project that was intended to improve water conservation in different countries to track the changes in the behaviour of the regional and local "systems" of water consumption using the methodology of outcome mapping. Study participant D co-designed and employed a series of observational techniques and surveys to document the behaviour of participants at a museum exhibit.

Some study participants are very deliberate in building the capacity of evaluation users to gather and analyze data. Study participant E described the strategy of their organization to build the skills and confidence of their clients by co-designing the instruments to get feedback on their work – regardless of whether the purpose is for developmental, formative or summative. In situations where the clients' capacity or confidence is low, his/her evaluation firm will assume initial responsibility for gathering and analyzing the data, but then encourage the evaluation user to assume a greater role in that process over time, providing "coaching" and "technical" support as needed.

(b) making sense of data

"Data is just data", reported study participant I, "it only turns into useful information and knowledge when you try to understand what it means for your program".

Several study participants stressed their role in helping evaluation users make sense of data. Study participant B, for example, recounts an occasion in which a group of managers were struggling to interpret data used to monitor the performance of their program: *The programs say that we have made a 30% gain. I respond by asking 30% of what – and they don't know.* Study participant F describes the challenge of making sense of statistical data about the school attendance rates of "kids":

> Measureable data does not speak for itself. It has to be interpreted within the context it is generated to have meaning. I am working with a program that is measuring a lot of things related to helping at-risk kids: e.g. their school attendance, their ability to secure housing, etc. Because the kids who come into the program are so unique, however, these statistics only make sense when you try to understand the significance of these measureable changes in the context of their story, for their perspective, and also using other non-measureable indicators of change. An attendance record of 50% at school may represent an amazing accomplishment for one kid and real step back for another one, particularly if the former is living on the street.

Developmental evaluators can help make sense of the cause and effect dynamics behind observed changes. Study participant G, for example, describes how his/her evaluation team uses contribution analysis to understand the role and 'value-added' of the many organizations and networks involved in creating new policies and programs to reduce poverty, a process which involves having multiple stakeholders share their perspective on the key contributions of various actors to changes in the local (e.g. a new social housing project, the reshaping of a provincial policy). Study participant E describes a similar process:

> We just did a survey with a large number of non-profits where 87% of respondents reported that they are tracking statistics on a regular basis. Some of them probably don't know why they are tracking this data any of them are unclear about how to transform it into something usable. We can try to help with that. For example, one of the organizations we work with keeps pretty good records of their meeting with targeted decision-makers (e.g. legislators, their staff, civil servants): e.g. this person, this date, this topic and a small narrative of what they discussed. They had 18 months of these logs. We analyzed these records together to identify which people they met with the most and then looked at "Lexus Nexus" to determine if they were using the key messages and, if so, how long after our meetings and ask the organization was getting a good enough return on investment by meeting with them.

Study participants confirmed that the process of making sense of data can be difficult. Study participant K reported "Sometimes the data is subtle, like the body language of a group of people in the room", and continued to argue that the role of the evaluator is to "make that kind of visceral data real" for people. This might include pointing out that some members of a management team do not appear to be comfortable with a recent decision to change the eligibility criteria for a leadership development program or that there appeared to be a difference of opinion about which strategy to pursue.

At other times, the volume of data is very large and confusing. Study participant I observed that emergent intervention tend to generate a lot of data (e.g. a lot of discussions about how to view the challenge, what directions to take, the effects of multiple probes and experiments) and it is easy to experience data-overload. He described the challenge as encouraging his clients to approach the challenge of sense making like a mystery, rather than a puzzle[30]. In a puzzle, he stated, there is a right answer and the more information you have, the easier it is to solve: for example, the more information we have on Osama Bin Laden's whereabouts, the easier it will be to find him. In a mystery, however, there is no obvious answer; e.g. what will happen in Iraq after the invasion? How is Enron reporting his financial health to the public? To work your way through a mystery, he argued, the answer is not more information, which may actually make matters worse, turning data into "noise" rather than helpful "signals". The more productive approach, he continued, is to focus on "sense-making" and help people better analyze the data that does exist, interpreting it from different perspectives and drawing working conclusions.

Yet at other times, the data are sensitive and difficult to use. Study participants B, L, N and R, for example, described situations in which organizations receiving grants from foundations reported that the actions of the organizations funding had a negative effect on their work. This included grantee perceptions that funders were imposing too many restrictions on the use of funds, were overly influencing the shape of the emerging intervention. Study participant I feels that the role of the evaluator is to "pull this data from the shadows" and "bring it into the

---

[30] This metaphor was popularized by Malcolm Gladwell in his book, What the Dog Saw and Other Adventures. 2009. Little Brown and Company. New York.

discussions" to explore what the implications were for the program and the working relationships between grantor and grantee. This required the ability and confidence to "speak truth to power".

Where the available data and its analysis are weak, evaluators and evaluation users have to be careful about the strength of the conclusions they draw. Study participant E described how subtle this process when trying to determine the role of advocacy coalitions in encouraging policy makers to adopt a particular perspective or position:

> I am always careful when drawing and framing conclusions. For example, if we know that on September 5th we met with a policy maker and on the 13th, he reiterated the key points of our meeting almost word-for-word, we can say that we likely influenced that person to move in his support from level x to level y. If we don't have a record of what was said in the meeting on the 5th, then we can't compare his comments on the 13th to the meeting discussion on the 5th. Maybe we made a difference, but we are not sure. Then it's safer to conclude that this person is moving in the right direction, but we don't claim that our efforts contributed to this.

Study participants also used different techniques to assist with "sense-making", including (a) multiple perspectives exercises which require people to assume the perspective of someone else on their team, (b) using "data scenarios" to encourage people to interpret different set of hypothetical data before sharing the research findings, and (c) role playing and simulations.

(c) data based decision-making

Most study participants reported that evaluators have to go beyond gathering data and facilitating a process of reflection by ensuring that evaluation users are making data-based decisions (A, B, C, D, E, F, G, , I, J, M, O , Q, R). "Hunches and reflections are not enough" when making program decisions exclaimed study participant O while study participant I argued. "The job of the developmental evaluator is to hold people's feet to the fire [of data]".

This is easier said than done. Several study participants described how even groups that embrace the idea of learning and reflective practice struggle make data central feature when it comes time to draw conclusions and make decisions. Study participant A described her

experience working with participants of an international agri-business venture who appeared eager to employ developmental evaluation from the very beginning of the project.

> They wanted to learn and were excited about gathering data and learning form the data. Previously they were learning but without hard evidence upon which to base (and define) their learning and best judgement… [My] role evolved over time. I tended to focus a lot more on getting the group to link their decisions more closely to evidence over time. We were at times making sweeping statements about things and making decisions and I wanted us to be clear about the data upon which we were doing this. So, while I tended to help create and lead 'reflective spaces', I felt more it increasingly important to 'bring clarity' to this space by bringing data, evidence and patterns into the discussion.

5.3.3 Tracking Developments

Patton and Gamble both stress the importance of 'tracking developments' of an emerging and ongoing intervention over time (Patton 1994, Gamble 2008). This includes "roads not taken, unintended consequences, incremental adjustments, tensions and sudden opportunities" (Gamble 2008). Documenting development serves a number of purposes: first, it provide innovators a record of why they made certain choices that may current affect current choices; next, it makes decision-making process more transparent, one way of ensuring accountability to learning; and, finally, it documents insights and lessons that may be useful for dissemination and the learning of other persons not involved in the intervention.

While only two study participants (A and I) identified tracking developments as a core role of the evaluator, several provided a variety of examples of tracking developments, even if they did not formerly identify this is as a role. Study participant C, for example, described how s/he and evaluation colleagues developed a unique way to describe and communicate a series of challenges they spotted in an organization that eventually came to be embraced as a regular form of tracking and reporting issues in the organization:

> We did some good things early on in a project that then bought us trust and space to do some crazy things. We wrote a "Houston, we have a problem!" memo for our client several years ago that worked out quite well. Now they want us to write four a year.

Other examples of tracking developments included (a) preparing and comparing visual diagrams of program models as they evolve over time (study participant B), (b) program reports that give an account of forks in the road and "roads not taken" (study participants J and A), (c) summaries of key learning's (study participant F), (d) memorandums confirming key moments in the evolutions in a program (study participant H), and (e) wicked questions faced by program designers (study participant R).

5.3.4 Emergent Practices or Questions

In addition to confirming the importance of framing issues, testing quick iterations, and tracking developments, study participants also surfaced a number of questions and debates related to the role and boundaries of developmental evaluator.

(a) to judge or not to judge

One of the distinctions that Patton makes between developmental and traditional evaluation relates to judgement. In a traditional summative evaluation, the role of the evaluator is to "judge the merit or worth" of an intervention, while in developmental evaluation his/her role is to assist intervention stakeholders develop an intervention that may or may eventually lead to a development that warrants a full-fledged summative evaluation.

Several study participants argued the stakeholders creating or restructuring interventions are making judgements about the merits or worth about specific elements of the emerging intervention – rather than a fully developed intervention -- all the time. This includes judging whether a particular way of framing a challenge or outcome is appropriate (study participant G); which strategy, a direction, or a fork in the road warrants attention over another (study participants E and J); or whether the initial effects and learning about an experiment or probe warrant continued effort (e.g. study participant R). Any time a group makes a decision or choice to proceed in one way and not another, they are in effect making a judgment call about the merit or worth of different options.

However, study participants have different opinions about the role of the evaluator in that process. Study participant G and I, for instance, reported that the evaluator should limit his/her

role to facilitating the process of judgement, jumping in only to point out when decision-makers fail to acknowledge when a critical judgement is being made and/or to ensure that they consider all the available data when making that judgement. Study participant B, on the other hand, felt the evaluator should "weigh in" with his/her perspective if s/he thought the judgement of the intervention stakeholders was not sufficiently robust and/of s/he would have made a different judgement:

> I go to their mechanism of change and their claim about how it is going to change some type of behaviour. If it's poorly conceived, then there is work to do. If well-conceived, and it's plausible, then I won't try to change it much (even if I don't like the program). So, my job is to get them to help them develop a plausible mechanism.

Study C went the furthest, arguing that evaluators have a core responsibility to offer judgement about the merit or worth of different parts of the emerging intervention. He went so far as to warn evaluators about the consequences of "being distracted" from that role:

> You need to know and feel confident about your primary purpose – to judge merit or worth. People may be surprised about this, but I am a Scriven'ite to my core: our job is to judge merit or worth of people's efforts. In jobs where people's roles are uncertain, and your role as an evaluator is swinging between active participant, organizational development work, facilitator, etc. it's very easy to lose this point. So, you have been confident to your work about the role of evaluation and what being evaluative is: that is, to judge.

The difference in practice by those interviewed appears to reflect a possible critique of developmental evaluation by other evaluators, which study participant B described as follows:

> The major opposition to DE tends to be about its emphasis on development and the apparent lack of judgement making. If that's the case, I think evaluation needs to grow up. We use the same tools in developmental and formative. Summative we are making judging about merit or, on formative keep eyes on merit or worth, eventually. Developmental evaluation does the same. I don't see the problem.31

At least one participant (E) was keen for Patton to share his perspective on the difference between the "big" judgement of the merit or worth of a program and the "smaller" judgements of

---

31 The author has not been able to find this critique in the literature on developmental evaluation. The study participants reported that this discussion has unfolded in his/her interactions with other evaluators.

the quality of developments that comprise an intervention, such as a new process for referring clients to a service.

(b) getting involved in the actual development

Several study participants identified instances in which they moved beyond relatively obvious evaluation activities and became actively involved in facilitating the development of the intervention.

This might include assisting a group overcome a barrier that gets surfaced during through evaluative activities. For example, study participant F described how s/he was involved in the early days of an effort to improve the school attendance of at-risk kids and chose to help the participating organization better elaborate and test how they would work together:

> The Executive Committee kept saying that they wanted to work at system level issues but in practice spent most of their time micro-managing program development. This was in part because the group had developed a few very general guiding principles, but they were not specific enough to really guide development – so the group didn't trust the operational committee to be able to move the initiative forward without the careful oversight of the Executive Committee. When I probed a bit to find out what would facilitate a hand-over of program design to the ops committee, I learned that the Executive would likely feel more comfortable if program parameters and principles were collaboratively developed, I facilitated this process which resulted in a document that outlines principles and parameters around four key areas: admission criteria, outreach/support, transitions/discharge, and recruiting. It was easy, productive and helpful piece of work. It was a DE 'deliverable' – to develop that part of the program and to helped us move the work – and DE work – along.

While evaluators may assume the role of facilitators in the design process, not everyone in the study feels that it is appropriate for evaluators to move beyond evaluation activities. One interviewee (C) reported that while s/he was very familiar with the pull to get directly involved in the day to day work of creating or managing an initiative, s/he felt that evaluators who did so ran the risk of eventually undermining their ability to play a productive evaluation role.

> Peter Block talks about this, the slide down the pole of a consultant -- eventually becoming a spare pair of hands -- and how difficult it is to climb up again. So, you have been confident about the role of evaluation and what being evaluative is.

Participant K also identified the challenge of not overstepping the boundaries between organizational development and developmental evaluation and admitting that s/he has stepped over it him/herself:

> We sit at the table, but we are not implementers. Where is the line? We need to guard against becoming vested as an implementer or involved in implementation. We need to be more explicit about that. […] I have crossed the line – usually one specific piece of it – but I am trained in this and so can pull myself back.

Participant G pointed out that Patton (2008) advocates for flexibility in roles to help with "reality testing" in his book *Utilization-Focused Evaluation*, but that he does not appear to address issues related to boundaries in his book *Developmental Evaluation.*

5.4 Accountability

Patton and Gamble are very clear about one dimension of accountability in developmental situations: whereas in traditional evaluations, accountability is directed towards external authorities and funders, in developmental evaluation, the primary accountability is internal to the innovator, the primary intended user of the evaluation process and results, and is rooted in his/her values and urge to make a difference.

The report from early adopters suggests that the current theory and practice of developmental evaluation is less clear regarding to whom the developmental evaluator is accountable, for what they are accountable, and how they demonstrate accountability.

5.4.1 Competing Paradigms

Study participant C summarized the confusion and/or tension felt and pointed out that this tension is rooted in a deeper tension about the paradigm, orientation or lens with which evaluators and evaluation users view the nature of evaluation work in emergent situations and used a framework developed by James Q. Wilson (1995) to describe challenges of accountability in the public sector.

The "Wilson Matrix" makes a distinction between four broad types of work, differentiated by the extent to which the outputs and outcomes of their efforts are observable or not. In production-oriented work, outputs and outcomes are observable: e.g. "the work of a police officer are radio calls answered, beats walked, tickets written, accidents investigated, arrests made while the outcomes are the changes – if any – in the level of safety, security, order, and amenity in the community" . In procedural work, the outputs of the work are visible, but the outcomes may not be: e.g. accounting where people are bound to follow strict operational guidelines in conflict zone. In craft-type work, the outputs are not observable, but the outcomes often are: e.g. coaching a sports team. Finally, in coping work, neither the outputs nor outcomes are observable: e.g. a social worker whose efforts to support a vulnerable family may be shaped by some minimum procedures, but whose activities are emergent and successes are difficult to detect.

Wilson argues that the concept of accountability plays out differently for each of these types of work. Accountability in production and procedural situations reflects a conventional understanding of the term: it requires someone to give an account of their actions, typically to a superior in a hierarchical relationship whose job it is to supervise their work. This type of accountability is straightforward to exercise because the outputs of work in both cases are relatively predictable, clear and visible and therefore lend themselves to formal reporting, monitoring and supervision.

Table 5-4 Paradigms of Work

	Observable Outcomes	
	Yes	No
Observable Outputs Yes	Production (e.g. passport provision)	Procedural (e.g. prison management)
Observable Outputs No	Craft (e.g. policing)	Coping (e.g. social working)

Source: Gregory (1995)

Conventional forms of accountability are difficult to employ in craft and coping work where the outputs of the work are difficult to anticipate in advance and often "highly dependant on the discretionary exercise of specialized knowledge".   Moreover, the outputs are often evident only in retrospect, once an outcome has been achieved, which in the case of coping work often remain elusive and sometimes illusory.   A social worker working to improve the interactions between a family in crisis, for example, must be highly responsive to the unique nature of that family, doing whatever is in his/her professional judgement at the time, and yet cannot be sure that even his/her best efforts will be sufficient to make much of a difference.

Because of their qualitatively different nature, Wilson argues that in craft and coping work, the legalistic concept of 'accountability to' should be replaced by the more appropriate 'subjective responsibility for'. Subjective responsibility is not externally imposed duty, but a

personal and professional sense of obligation to significant others for acting with "prudence, good judgement or moral probity" and guided by a "strong sense of shared mission, commitment to professional norms, standards and values, and above all on maintaining high levels of mutual trust and respect".

Many study participants describe developmental evaluation as a version of craft or coping work described by Wilson. Study participants B, I, F and R describe how the exact nature of meaningful developmental evaluation emerges only after a period of time while study participants F, K and L argue that when they do emerge, much of it is "somewhat invisible" and difficult to describe. As such, some participants are not convinced that reporting of deliverables is the only or best mechanism to ensure accountability. Study participant C, for example, notes, "A report seems an odd thing in DE." while study participant F argues, "It is easy to produce reports to demonstrate something but that may not be the most useful approach". Study participants J and M argued that evaluators are responsible for principles, ethics and guidelines of the evaluation profession.

5.4.2 Emerging Questions and Issues

(a) accountable to whom?

Regardless of whether they felt more comfortable with a production/procedural or craft/coping paradigm, study participants struggled to determine the primary intended user in a developmental situation.

(i) contractors

Some study participants felt that they are accountable to the evaluation users with whom they contract. As study participant M described his/her approach:

> This is not so 'vexing' for me. As a contract evaluator, my accountability is always clear – it's to the person with whom I sign the contract (innovator or funder) to do whatever we agree I should do. It may change frequently, there may be scope creep, a shift in how much I do voluntarily and off the clock or even ask them to pay for. We always have to (re)negotiate this as it unfolds - but it's always clear.

Study participants H and R agreed and argued that even though they were "technically" accountable to a board of directors and senior management team of the organization overseeing the intervention, in practice, evaluators are most accountable to the person responsible for engaging them to provide evaluation support.

Study participants A E, F, I, L, M, S and R described the difficulties that emerge when one of the users of a developmental evaluation is a funding organization underwriting the costs of the intervention-in-development. While Patton (Westley et al. 2006) argues that in developmental situations funders are ideally equal members of the team that is developing the intervention and therefore should not have any more power or influence in the intervention and evaluation than other members of the team, in practice the reality is more complex.

Study participant B argues that funder supported evaluations immediately creates awkward dynamics for the evaluation work. Study participant B notes:

> I think it matters who brings you to the evaluation. For example, if I was brought in by the funder, I carry the authority of the funder. I eventually have to get free of that authority because if I don't, I will be telling them what I want to see and/or they may be less open with me. It's an awkward position. I generally emphasize that I want to talk to you [the grantees] without this going back to the funder and that my job is to work with you and leave the benefits here. Once that is established, I can provide useful information and will bring it to the donor, when we have some understanding of this.

Study participants E, F and R described how it can take a great deal of time for the evaluator to break free of this authority and establish a trustful relationship with the primary evaluation users. Study participant F, for example, reported that in one instance it took an entire year before program designers appeared to accept their independence of the funder. Study participant E argued that that this process of acceptance can be accelerated by involving the innovators as early on in the process as possible:

> When the funder is paying for it, the grantee has to be involved in the very beginning – even before the contract is signed – or else it does not work well. We have experiences where the contract was signed before and the funder introduced us as the evaluator. This can be uncomfortable for them and I have lost months of time working on building up the relationship and trust with the intended user.

Even when the evaluation users have been involved in the evaluation process from the very beginning, the evaluator is still apt to become tangled up in power dynamics between funders and grantees, something study participant R felt was unavoidable:

> The overlap between the [evaluation users] from the sites and the funder was tricky. We were the ears on the ground for nine months. We head things – sensitive things – that funders would not normally hear from site evaluation and perhaps were not ready to hear. There was power dynamics involved. If you are asking grantees to be critical about how things are unfolding and they tell you that it includes the funders work, what dynamic does it stir up? Are the [funders] really ready for developmental evaluation?

Whether or not funders were ready for developmental evaluation, study participants B, F, I, L, and R described situations where they met with funders to describe how their activities or expectations were negatively affecting – or perceived to be negatively affecting -- an emergent intervention, a role that study participant I describes as "speaking truth to power".

Should evaluators resign themselves to simply trying to navigate funder-grantee dynamics, or, should they try to change them? Study participants described at least two different options. Study participant A argued that where funders are unable or unwilling to participate as team, they should be kept at arm's length from the day-to-day decisions about the intervention-in-development and be considered as a secondary intended user, rather than a primary intended user. While they might be approached to become involved if the evaluative process turned up an issue that required their attention, they would

Other study participants preferred to encourage funders to become more involved in the work of the grantees. Study participant B and Q called this approach "relationship grant making" inspired by the practice of relationship banking in the private sector. Relationship grant making may improve the social capital or trustful relationships between grantors-grantees as well as embedded the grant officers in the emergent intervention so that they are active and productive members of the team managing an intervention-in-development.

(ii) working at different scales

Study participant D, F, L, I, J, L, M, N and R described the challenge of being accountable to evaluation users operating at different scales of the intervention(s). When describing her experience with assisting staff and the management of the science museum develop exhibits to encourage innovative thinking, study participant D reported that she struggled to find the time to fully engage the higher leadership of the organization in the process. As a result, she felt the effects of the evaluation were largely confined to the staff working at the level of individual projects. .

(iii) turnover in users

Several study participants G, O, R reported how the coming and going of evaluation users in emergent interventions complicated the challenge of how evaluation users would come and go with the evolution of the emergent intervention. Study participant R described a situation how early on in the project, their primary accountability for her work was to the senior manager who brought her into fast growing project on leadership development for young women. This changed with the arrival of several new managers hired to deal with the explosive growth of the program, requiring him/her to establish entirely new relationships.

(b) accountable for what?

For what is an evaluator accountable for in developmental evaluation? What is success in developmental evaluation?

Study participant I and F argued that the primary measure of success was the extent to which the primary intended users "learned" something through rigorous thinking and data. This could be learning about the problem they are trying to address, the emerging intervention and/or the landscape in which it is unfolding.

Some participants felt that "use" of evaluation findings was the key indicator of success in developmental evaluation. Study participant B argued that learning was not enough if it does not shape the day-to-day decisions of evaluation users. S/he went out to argue that helping policy makers, managers and front line workers practically link or integrate learning, reflection and thinking into their day-to-day work was a challenge for the field of evaluation in general:

The evaluation world has not yet fully thought through the idea of evaluative thinking. Evaluative thinking is thinking; management is about the use of thinking. You would be hard pressed to find a hand full people who practice evaluative thinking and doing. We are trying to get people to think against their own interest and I don't think we appreciate how difficult it is. But it's necessary because, there are trillions of dollars at stake.

Study participant I shared a similar opinion and described how the extent to which evaluation use learning and use as criterion for success depended in part on their confidence and trust in the evaluator:

How successful this is depends on the conditions for good DE. If and when we chase a rabbit down a hole, oops, we both risked. If the ecology is good, then we call it shared learning. If the ecology is bad – for example, with low trust – then they might say that you – the evaluator -- wasted our time.

Is a developmental process does not yield a concrete and practical intervention – and the developmental evaluation supporting the work – a failure? Study participants G and I described instances in which their work did not result in a new intervention or tangible shifts in an emergent one and several others referred to occasions of work that "went no where". Study participant K described what she felt was a paradoxical nature of success in emergent situations and developmental evaluations:

On the surface its "deliverables" but deliverables are not at the core of it. There are deliverable and there are deliverables. It's to help develop something with good data, rigorously, evaluatively. [But], if nothing is developed, is the effort – including developmental evaluation – a failure? It would be a shame, but it may that the initial thrust was not well founded. On the other hand, it would not be a disappointment if it meant that people saved a lot of time and money. People can sometimes be pretty rigid about pursuing one idea or way of achieving something. The evaluation can surface that the assumptions underlying the idea don't hold true.

Finally, study participant E stressed a "long view" of evaluation and an emphasis on "process use" that is, getting evaluation users to think evaluatively.[32] All "bursts of evaluation come and go", he argued, but the need for organizations and groups to develop a culture and capacity for evaluation and learning is never ending. As a result, the ultimate measure of any

---

[32] The concept of process use has been popularized by Patton (2006) in his writings on utilization-focused evaluation.

developmental evaluation effort should be the extent to which evaluation users – either in a collaboration, project or organization – are building a stronger commitment to, and capability to use, evaluative thinking in their day-to-day work.

(c) accountable for how?

Only one study participant made it a point to describe that they were accountable for the way in which they provide evaluation support. Study participant J emphasized that the evaluators should be diligent in following the ethical and professional guidelines established by their professional societies. This included the Canadian Evaluation Society for ethical conduct which outlines eleven points in the areas of competence, integrity and accountability.

<div style="border:1px solid black; padding:10px;">

Table 5-5 CES Guidelines for Ethical Conduct
http://www.evaluationcanada.ca/site.cgi?s=5&ss=4&_lang=EN

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.
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.
Evaluators should be responsible for the clear, accurate, and fair, written and/or oral presentation of study findings and limitations, and recommendations.
Evaluators should be responsible in their fiscal decision-making so that expenditures are accounted for and clients receive good value for their dollars.
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.

</div>

5.5 Situational Recognition

What processes and criteria are evaluators using to determine the extent to which something is appropriate for developmental evaluation? While neither Patton nor Gamble have described this process in much detail, study participants have devised their own approaches.

5.5.1 The Process

Approximately one-third of study participants (A, E, J, K, M, Q) reported that they employed semi-structured or very structured processes to determine the purpose of the evaluation (e.g. summative, formative, developmental), to surface the possible implications for evaluation activity (e.g. questions, preferred methods), and to help them make a decision about whether or not to proceed. They alternatively described this process as the "diagnostic", "discovery", "exploration" or the "scoping" phase of an evaluation. All of them employed generic "discovery" processes, rather than ones that specifically designed for developmental evaluation.

5.5.2 Aides for Action

Study participants differed to the extent to which they employed reported using tools or techniques and/or their "instincts" in their efforts to determine the suitability of employing developmental evaluation in a given situation.

Study participants E, F, M, Q and P reported using various frameworks and lenses to determine the purpose of the evaluation. Study participant P, for example, uses a "simple-complex acid test" to determine the degree of certainty:

> I ask my client if she is confident she knows the relations of cause and effect between what she proposes to do and what the results will be. If she does know, then it is a "simple" situation. She is challenged to do the right things right in order to bring about change. And, DE is not for her, although she would benefit from a formative evaluation mid-way through to make adjustments to keep her on track. If, however, she cannot say with certainty what she will achieve, but is confident that by doing what feels is right she will find the way forward to the change she wants to see, her challenge is "complex". She does not know the relations of cause and effect. This situation is ripe for a developmental evaluator to help her identify and understand, in real time, her results and how she contributed to them.

Some of the tools are more elaborate. Study participant E, for example, uses a checklist with ten questions to assess the readiness of a group to get involved in an evaluation and to surface issues that should be addressed in the evaluation. Study participant M has developed a fairly comprehensive planning guide to help evaluators and evaluation users prepare to evaluate efforts aimed at systems change.

> I created a planning guide, a book with worksheets [that] includes a brief on evaluating systems change. It's based on the ideas that you need to understand the situation and you need to articulate the intervention, the governance structure, their assumptions about how to create change with theory of action, their hoped-for outcomes, how predictable and the diversity they are, etc. Those are not different what you normally do [in traditional evaluation], but the level of complexity is different. It helps people think through the purpose of an evaluation.

In contrast to those participants who used tangible tools and techniques, other study participants emphasized using more inductive and intuitive process to better understand the work of prospective evaluation users and to determine the extent to which a situation was developmental. This involved looking at how prospective evaluation users speak or write about their work (See Table 5-3).

Table 5-6
"Soft" Approaches to Detecting Developmental Situations

I know it when I see it. Or better yet, I intuitively know when it isn't developmental. Our work is so much focused on a developmental approach that I am alerted to a situation when it isn't. Study Participant K.

How do I know? I read documents, I talk to people and I try to understand the project story. We try to summarize this and to articulate their theory of change: i.e. what they are doing, why, what do you want to change, the context in which they are working, to what extent have they accounted for dynamic factors in the project. I am big on how people open sentences: e.g. "I am not sure"; "What do you think?"; Could we?" I look for grey areas, soft – rather than hard – language, quantity of text, etc. Study participant L

I also look at the way they tell their story: are they in the story and shaping it, or is 'it' happening to them? I get some of this language and these ideas from Ron Short, author of Learning in Relationships and other stuff on emotional intelligence. Study Participant O.

5.5.3 Biases

Several interviewees pointed out that some evaluators - and possibly evaluation users – demonstrate a bias for seeing a developmental nature and a need for developmental evaluation even in situations. Study participant K notes:

> I actually think most things are more developmental than people realize. We are not as sure about how something work or the environment is changing more quickly than we realize. I think most things are more developmental than not but we choose not to look close or deep enough to see it.

This bias does not sit well with some of the interviewees who acknowledge that it exists and feel that that it undermines the objectivity of the evaluation process and the professionalism of the evaluation profession. Study participant M noted:

> I have an argument with people who see the whole world as developmental and adaptive. Sometimes something is and sometimes it is not. Let's assume that we don't know. This is particularly true in government programs which usually have aspects that are simple, complicated with some degree of complex.

Study participant K described a similar concern about the bias, but argued that it was not problematic if the evaluator and evaluation users were conscious of it and made a disciplined effort to work in situations that were obviously emergent. She noted: "I think [traditional evaluation assignments] used to be 70% of my work and now it's more like 50%. The shift is because I am focusing more on where my interests lie so seek out developmental situations."

5.6 Methods

Study participants were not asked directly about their experience and reflections on methodologies in developmental evaluation, yet their responses to the interview questions surfaced a number of features that may be important in any developmental evaluation effort.

5.6.1 Common Features

(a) contingency-based

Study participants appear to agree with Patton's argument that there is not standard or unique set of methodologies employed in developmental evaluation and that the appropriate methods are situation-specific. When the researcher pressed several participants to describe typical methodologies and techniques, they responded that methods should be customized to reflect the unique context of each evaluation. For example, study participant B responded:

> You have to be competent and comfortable with methods. However, it's not about a single set of methods. DE – or any evaluation - is not driven by techniques, method or a certain type of rigour. It's not about us and our ideal methods – it's about them, the clients or users, and what the situation requires.

While the general sentiment among study participants was that developmental evaluation was "methodological agnostic" (Study Participant I), some did profess a bias for qualitative methods. Study participant A, for example, revealed that s/he had a strong preference for qualitative methodologies because s/he felt that they better suited emergent situations and because s/he had little training in quantitative methods. Study participants D and H reported that associated quantitative methods – with their emphasis on measuring the effects of interventions -- more with summative evaluation.

(b) formulating evaluative questions

Study participants A, C, D, R, E, O, and I argued that surfacing, developing and pursuing productive questions was a central feature in developmental evaluation. This includes questions that are typical of traditional evaluation (e.g. what are the effects of our work? How can we improve this model?). It also include questions that are "tough", "tricky", "wicked" and "stupid" that aim to encourage participants to think about the nature of the problems they are trying to address and the deeper assumptions behind their emerging interventions. Study participant C reported:

We ask 'stupid questions' to expose assumptions and help test and set boundaries. What's worthwhile and not? What are good and not-so-good directions to take? We bring question posing skills to the work.

Posing questions is one thing. Answering them is quite another. Study participant C went on to express their concern that the emphasis on reflective practice and evaluative thinking do not emphasize the use of the data in practice:

I am getting tired of the focus on questions. In fact, it's the answers to those questions that we are seeking. So, yes, questions but let's not stop there: that's just the beginning. Evaluators need to help get answer to those questions so people can make decisions.

Study participants B and R shared the perspective, arguing that the answers to the questions ultimately need to help evaluation users make practical decisions about how to develop or manage their work.

(c) primary and secondary data

Developmental evaluators may often draw on both primary and secondary data to inform their work. The use of multiple methods and data sources is typically considered a feature of a robust evaluation. Several study participants reported that it was particularly important in emergent situations where the ability to pull together primary data is limited by the challenge of trying to track multiple, short term probes and experiments that may generate only weak effects – if any at all.

Study participant F illustrates his/her extensive use of both types of data to inform the development of an intervention to improve health in a targeted demographic group:

For example, I used a lot of research to understand which risk factors – modifiable risk factors –would have the biggest impact on a particular health issue. We then drew on more on environmental scans, literature review, key informant interviews, demographic data, etc. to further inform and test iterations of the model.

Study participant G described how the evaluation team supporting a network of urban organizations working to convince large employers – particularly municipal governments – to pay their employees a "living wage" encouraged them to use the large pool of secondary

evidence on the effects of such policies on employees, employers and municipal budgets rather than assume that only locally generated data was suitable.

(d) user friendly language, concepts and measures

The emphasis on user-friendly language and concepts was the third common characteristic mentioned. Study participants A, E, J and K argued, the evaluator had to communicate ideas and data in a way that was most likely to be used by the people making decisions about the intervention.

One strategy is to employ the language and concepts of the people who use the information. Study participant B described adopting a measure or indicator that made sense to participants of a fisheries project:

> Often, local people have a lot of information but because of their knowledge base or logic, they have the information in a different form [than used by evaluators or evaluation users]. For example, in one fisheries project on the east coast, local folks used a metric called "catch and effort", i.e. how much time does it take to fill our table with food for a meal? For example, they might say, "It used to take six hours and now it takes three". In that case, do we – coming in from the outside – need anything else or much else? I know that we used huge amounts of time in journals debating this and I thought this simple metric was perfectly fine.

Study participant M emphasized that building common language and measures take time, but the resulting buy-in by the evaluation users may be worth it. They illustrated this with the following example:

> With any evaluation, you need to get their buy-in that it's useful for them. Maybe ten or twenty years ago, you could get someone to collect information for the sake of collecting it. Now, there is so much work that goes into making this happen: developing a theory, measures, confirming the value of information. For example, we worked 500 grantees for a year to develop measures that made sense for them: people at the clinic level have supremely relevant and useful needs, these stayed at the final set.

Some study participants reported using visual aides such as diagrams, charts and mind-maps to make evaluative data and discussions user-friendly. Study participant B, for example, described how she would encourage their clients to use large sheets of butcher paper to draw

their current understandings of the interconnected problems they were trying to address as well as the characteristics of their emerging intervention. She would then compare the various snapshots over time to get a sense of the evolution of their work.

(e) participatory

Several study participants identified "participatory processes" were an important methodological feature of developmental evaluation. Study participants A, D, E and K reported that having evaluation users involved in developing evaluative questions as well as gathering, analyzing and drawing conclusions about that data to inform their decision-making is important for a variety of reasons. This includes improving the probabilities of more accurate and timely information; building the ownership of the users of the process and findings; and, enhancing the capacity of the group to do evaluative work in the future.

(f) adaptive

The fifth major characteristic of DE methods identified by interviewees is an adaptive approach to methodology. Several study participants described experiences in which the methodologies they employed evolved over time. When asked about the evolution of methodologies and roles, for example, study participant E reported:

> I can't think of many examples where it does not evolve. Perhaps short projects – e.g. 4 months to 6 months -- stay the same. But for instance, in the project that lasted five years, the role and methodology changed in response to changing contexts, [the clients'] strategies, their leadership (including my key contact), and their capacities.

> Sometimes the evolution can be quite sudden.

> In one quarter, the work on network impact was the thing to do but we changed the next quarter because we decided to run an annual retreat.

Other study participants were careful to point out that adaptive design does not always mean a wholesale restructuring of the methodology. Study participants E, J and K, for instance, described instances in which some of the information evaluation users required on a consistent

118

basis was relatively clear and therefore it was "fairly easy" to design and implement relatively stable processes and systems for gathering and analyzing the data.

5.6.2 Emerging Questions & Issues

(a) rapid or timely feedback?

Study participants B, K and Q argue that "rapid feedback" is another characteristic of developmental evaluation because evaluation users work quickly when choosing which avenues appear to be dead ends and which show promise. Study participant B reports:

> My job is to help them focus and help develop a way – using a short information cycle as possible – to determine how things are going. It feeds back into the management of their project and how they use the information.

While rapid feedback may be a typical methodological feature in developmental evaluation, it may not be required in all circumstances. Study participants I and N, for instance, described a project in which they were both active at different times which involved a team working on an ambitious new strategy to improve the quality of persons with disabilities. While the group spent a great deal of time discussing the nature of the challenge and different ways they might proceed, much of the time very little happened and when it did, there was a short lag time between their activities and results. As a consequence, study participant I argued that developmental evaluation methods should focus on providing timely feedback, which may or may not need to be rapid.

(b) balancing reliability and validity

Several study participants described the typical evaluation challenge of balancing reliable and validity in their research.

Developing reliable methodologies is often very difficult to realize in developmental situations where the evaluation questions of the primary users – and therefore often the research methodology – are apt to change quickly in response to new learning, changes in stakeholders and/or shifts in the intervention. Study participant Q, for example, described how the evaluation

priorities of a national organization stewarding leadership programs for girls changed from quarter to quarter.

The tension between reliability and validity can be particularly vexing for evaluators steeped in the tradition of conventional social science research. Study participant C, for example, described a situation where evaluation researchers in a large project got "caught up" with carrying out a rigorous analysis of data that did not reflect the questions of the evaluation users:

> I am seeing my colleagues slide into debates about collecting and coding material. I am asking how it's going to be used. We just recoded the data and realized that we did not ask the kind of questions our client was interested in. My colleagues are willing to experiment at the edges but they are classic social scientists when they perceive that the job is underway. They are technically very good at it, meticulously rigorous, but I can see an adaptability issue coming down like a train.

In the experience of study participant E, the struggle that experienced researchers have with developing and adapting imperfectly reliable research instruments may be part of a broader pattern rather than an exception:

> I am coach to several local evaluators in the region. A lot of them have been in the field for 20 years. I find that evaluators with 8-15 years of experience are most comfortable with developmental evaluation and its adaptive approach. Those with more experience with tried and true methods tend to worry about methodological rigor and often don't like changing instruments because it interferes with that rigor.

Even when evaluators feel comfortable with their ability to manage the tension between reliability and validity, evaluation users may be keen to favour reliability over validity. Study participant D, for example, described how the people creating a new program to increase the innovative behaviour of museum visitors through interactive exhibits wanted a more rigorous approach to tracking the effects of their efforts mid-way through the development of the program:

> They wanted rigorous data about the behaviours of the visitors so that it can be published. So they had a lot of concerns about sample size, methods, etc. that are accepted as minimum requirements for reliable data. They wanted to

demonstrate the value of the Center – or the evidence of some effect – to the science world. There was one process with a couple of parallel purposes underlying the evaluative work as it was unfolding. This sidetracked us a bit and took some resources away from other aspects of DE.

Study participant E expressed confidence about his/her ability to balance the demands of reliability and validity when she noted "Being flexible should not sacrifice gathering and using good data". She went on to describe the importance of understanding the limits of research instruments as well as being careful about drawing conclusions that accurately reflect the strength of the data and analysis they produce:

> I feel comfortable with weak and ambiguous data when I can stand behind the instrument, when I helped design it and know what it can and can't do. I am less confident when the data seems to say something, but I don't understand the instrument that produced it.

> I am always careful when drawing and framing conclusions. For example, if we know that on September 5th we met with a policy maker and on the 13th, he reiterated the key points of our meeting almost word for word; we can say that we likely influenced that person to move in his support from level x to level y. If we don't have a record of what was said in the meeting on the 5th, then we can't compare his comments on the 13th to the meeting discussion on the 5th. Maybe we made a difference, but we are not sure. Then it's safer to conclude that this person is moving in the right direction, but we don't claim that our efforts contributed to this.

The tension between reliability and validity requires evaluators to be confident and disciplined when drawing conclusions and making claims. Study participant G, for example, describes the pressure his evaluation team feels from some of the members of a national network of collaborations working to reduce poverty to report that their efforts were largely responsible for observed improvements in thousands of low income households, despite the fact that data required to make such a claim does not exist.

(c) a bias for quantitative data

Several study participants (D, F, and G) identified the challenge of working with evaluation users who feel quantitative data is superior to qualitative data. This may include the people making the day-to-day decisions about the intervention, other evaluators or researchers, or even secondary users, such as program funders. While a measurement bias may be a subset of

the larger challenge of balancing reliability and validity, it is sufficient unique to present on its own.

Study participant E, for example, describes how the efforts to develop for standardized and quantitative measures for a program for at-risk kids weakened the overall evaluation effort:

> Metrics are important – we need ways to determine whether or not we are making progress. However, not everything can be measured, and the more complex the initiative, the more difficult it is to develop meaningful indicators. In the face of this challenge, I think we often compromise significance, focusing our efforts on what can be measured quantitatively rather than investing in processes that would yield more meaningful (albeit more messy) data.

> We also get into trouble because people want standardized measures. In an educational program for at-risk kids that I've been working on, the entry points for the clients are so different that it's difficult to develop standard ideas of success: for one student, the ability to transition to a community school and maintain regular attendance for six months would be a good measure of success; for the street-entrenched youth with significant mental health issues, transition to a regular school may never be possible. The program can still have a very positive impact on that young person, but our measures would never capture that.

Not everyone feels constrained by a bias for quantitative data. Study B, a seasoned evaluator, stressed that while she felt strongly that the preferences of evaluation users was central, there was a way to employ both quantitative and qualitative methods. She confirmed that she was comfortable with a range of methodologies, admitting his own preference for surveys and interviews, but ultimately felt that intended users should shape which techniques were used as well: "At the end of the day, you can get "Bar Charts" from either method [i.e. qualitative and quantitative] if your client wants it!"

5.7 Adaptive Design and Budgeting

How does an evaluator and evaluation user(s) develop an evaluation plan and budget for evaluation situations that can be highly emergent and uncertain? Study participants described four broad approaches and raised questions related to how to demonstrate and cost out the full range of evaluation activities and their value.

5.7.1 Options

(a) open-ended design

Several study participants describe an approach to planning and budgeting for evaluation work that study participant I described as open-ended. This involves developing a 'roughly right' direction and set of parameters to guide the evaluative work and then allowing it to unfold organically, with the evaluator and evaluation periodically coming together to decide what areas of work deserve attention.

> Sometimes we settle on broad directions and work relationally as the work emerges: e.g. [with one organization], we had a highly responsive, more of retainer, kind of relationship. In a good partnership, people are able to say "I wonder if this urgent item is less important and less urgent than that one". I might develop a scope of work that sets all of this out – more often than not – but it's not always the first thing.

An open-ended approach requires an evaluation budget that may be fixed in terms of the amount of resources available but open-ended in the way they are used (B, I, K, M). This allows the evaluator and evaluation user to agree on specific evaluation activities and their cost implications as they emerge, accounting for them and releasing payments after they have been completed. Study participant B, for example, describes his approach to developing an umbrella contract in emergent situations:

> I need to know that I am not going to be locked into a deliverable structure to which my resources are tied. If so, this represents a lack of flexibility. The user has to bring to the table an adaptable approach in the use of the resources. This is not so much the total dollars - these are often set – but in how they are used. I often use a technique called "draw down": a general pool of funds that can be employed for the evaluation but for which we talk before we use any of it. This helps mitigates risk. So for example, if I say we'll stay under $250,000 for x amount of time, for what I do, that is fine. The $250,000 may last two years or we may only use my $150,000 in that time. I sent in quarterly reports that describe what's going on and if we need to draw down some more. Then we have a conversation to see if we can agree on that.

Study participants argue that an open-ended approach requires a great deal of trust between the evaluator and the persons responsible for allocating resources (B, I, K, G). Without it, both sides will feel pressure to develop a fairly precise evaluation plan and budget and

emphasize close implementation to both as the evaluation activities unfold. This approach can reduce the responsiveness and flexibility of evaluation activities. It can also be more expensive if the evaluators and those commissioning their work are working with little trust. Study participant I uses a construction analogy to illustrate the point:

> The advice I got from my Dad on construction projects is that when you hire someone and say you need a firm quote, they build in all the unknowns and the ambiguities into the quote, which will usually be higher than normal. If you approach someone and describe what you hope to build, to ask him to give you a ball park cost and let him simply start building it and adapting as you go within that limit, it often ends up to be cheaper in the long run. In evaluation, when you don't have trust, you tend to over design and be deliverable oriented. When you have trust, it's generally process and ultimately outcome oriented and easier to deal with uncertainty and emergence.

This trust goes both ways: two participants (B, I) reported finishing development evaluation assignments under budget and returning funds to happily surprised clients.

(b) start with something small and tangible

Another option for developing plans and budgets for emergent situations is to start with a small set of evaluation activities that address an immediate and/or obvious set of evaluation questions and then expand the evaluation work over time. This allows the evaluator and evaluation users to plan for and budget a relatively concrete and manageable size of work in way that builds capacity and trust for further evaluative work.

Sometimes starting small begins by evaluating only one part of the emergent work. Study participant E described how his/her evaluation organization typically tested things out with clients by constructing a five-day small pilot in which they scoped, planned and implemented evaluative activities for a small area of work. He illustrated this by describing one part of the pilot with a group trying to get their message out in the media:

> For example, for organizations that work with the media, we might develop a media scorecard by retrospectively tracking key words related to their policy interest (e.g. maternal health or food security") in "Lexus Nexus" over a period of time and analyze how frequently the key words showed up, who used them, tracking this over time. This creates a baseline for "in-the-moment' monthly or

quarterly tracking of the media attention on this key issue in the future and gives them a quick taste of evaluation.

Similarly, study participant R used a prototype approach in the early days of evaluating a leadership program for young women:

> The DE work was done in baby-steps, doing little things first. People get to experience it, get better, gain confidence and want to try it out in other situations. Baby steps are good because [developmental evaluation] methods are process oriented, conceptual at times, and it's not for everyone. Many staff, board or management team they come from different levels to deal with conceptual stuff (e.g. complex systems, network theory) – not everyone is ready for this.

At other times, starting small can focus on a broad and narrow evaluative look of the entire emergent intervention, a strategy study participant I called "developmental evaluation light". Study participant C, for example, prefers to organize his work into "two parts, two jobs, two prices". In the first part, he works with the evaluation user to create a "mini-evaluation" or "fractal" of a potentially larger evaluative effort. The benefits of the mini-evaluation, he argues, is that not only is it relatively easy to plan and budget for, but it also allows the evaluator and client a hands-on approach to scoping out the parameters and value of a larger evaluative effort.

(c) plan and adjust the plan

Some study participants describe examples of attempts to develop a fairly comprehensive and detailed evaluation plan up front and then monitoring it closely so that it can be changed to reflect the evolving nature of the emergent intervention and evaluative pressures that it entails. Study participant R, for example, describes how she and her evaluation client continually reviewed and adjusted an evaluation plan that was originally designed to answer two to three evaluation questions:

> We did quarterly check-ins and work plan and budget updates. They were in the middle of massive ballooning phase – new money and new staff and quickly growing organization. It took me two quarters to figure out that this was important to do: we were learning how to do this. To say you are going to generate rigorous data is one thing then you need to have to learn how to do it and we had to make changes quarter by quarter.

This approach may be rigorous but it can also be expensive. It is often time consuming to find more than a few tangible and stable things around which to plan evaluation activity when few such opportunities exist. Study participant B, F and R reported examples where it was six months to a year before a clear pattern of evaluation questions and activities emerged. Furthermore, if and when a detailed evaluation plan is developed, the time and energy to monitor and upgrade them in equal detail is often exhaustive.

(d) patch planning and budgeting

Where some of the evaluation work is both emergent and straightforward, evaluators and evaluation users can develop a plan that is clear and fixed in some areas and open-ended in others. Study participant K, for example, argued that since much of her work included a combination of certain and uncertain elements, she often develops a plan that incorporates elements of both:

> As far as planning goes, certain pieces of work are pretty clear and stable: e.g. people need a certain piece of information and so we put systems in place that can get that information but can also accommodate and hold the changes. This means that while the content of the information may change, the processes are in place to get the content largely remain (e.g. quarterly reflection sessions, regular surveys with different contents, story collections). Other pieces of work are not so clear and we have to adapt over time.

There is variety of ways to budget for patch evaluation. Study participant K, for example, simply draws done a proportion of the budget on a regular basis (e.g. monthly, quarterly) regardless of the ebb and flow of evaluation activity, hoping that in the end, "everything comes out in the wash" but that sometimes s/he and her colleagues will put in effort for which they are not reimbursed. Study participant E, on the other hand, develops different "buckets" of activities and budgets which allows for 'fixed budgets and timelines' for stable and predictable areas of work and more 'open-ended' buckets for emergent areas.

5.7.2 Emerging Issues

(a) identifying primary and secondary users

While most study participants appeared to agree on the principle that the primary intended users of developmental evaluation should be the people who made the day-to-day decisions about the emerging intervention, several participants described periodic difficulties in determining who precisely are – or should be – primary and secondary users.

Sometimes the *real* decision-makers for an intervention are not involved in the formal structures and processes of decision-making but rather influence them from a distance. Study participant E described how it was sometimes difficult to determine if the organizational representatives at advocacy coalitions were truly the primary intended users of the evaluation information or perhaps only a liaison for decision-makers in their organization not attending meetings. Study participant A described her challenge of determining whether the funders of many international development projects were primary or secondary users given the varying degrees of their involvement and influence on the day to day decisions of the projects.

Primary and secondary users are apt to evolve over time and in the process reshape the evaluation. For instance, study participant D described a situation where philanthropic funders that had provided "patient funding from a distance" to a non-profit group suddenly became active in the project after two years when they become concerned that "nothing was being developed". Their rapid shift from secondary to primary evaluation shifted the emphasis of the evaluation from learning to development and then to accountability and outcomes within a real short period of time.

(b) disagreement among users

Another challenge experienced by study participants who work with evaluation is the struggle to agree on the priority evaluation questions and evaluation design.

Sometimes, evaluators try to facilitate and/or negotiate a shared agreement on the questions to be pursued and characteristics of the investigation. Study participant O, for example,

described how he typically invests a great deal of time in the exploration phase to determine the evaluation users, their expectations and preferences in an effort to determine if there is a core set of questions which all stakeholders agree should be pursued. If and when an authentic agreement is not possible, he often declines the assignment.

At other times, when getting agreement is not possible and the resources are available, evaluators choose to pursue multiple lines of investigation concurrently. For instance, study participant Q described a situation in which the funding representatives on a steering committee overseeing the evaluation of a cross site community mobilization and intercultural integration program were interested in learning about the extent to which the programs were working or not while the representatives for program grantees were more interested in the extent to which capacity of the local agencies to develop and evaluate mobilization activities was strengthened, the official goal of the evaluation. While she was able to pursue two lines of questions concurrently, she described the process as "tricky", "messy" and periodically "tension-filled".

(c) budgeting for invisible work

Some study participants described a variety of budgeting – and being compensated – for evaluation work that they felt was unique to developmental evaluation: long start-up times, invisible work, and the comparatively high transaction costs of adapting plans.

The emergent nature of emergent interventions means that it can take some time for concrete evaluation activities to become clear. Study participants I, F and R, for example, described situations where it took several months and even a year of interacting with evaluation users to understand their work, develop relationships, and carry out some meaningful evaluation activities. Study participant B noted:

> I think the stuff I "deliver" generally gets clarified after a year or more of work:
> it's sometimes impossible to know what that is any earlier. It has to be this way
> with inductive processes: things become a little clearer once we start something
> and see how things are going.

While this start-up time is short on evaluation activities, it still requires evaluators to invest time and energy participating in the emergent process, which presents a challenge in terms of predicting workload and evaluation budget.

Even once evaluative activities are underway, the outputs of that activity are not always tangible. Study participant F describes the challenge of making some types of developmental evaluation visible:

> One of the difficult parts of DE is that so much of the work happens behind the scenes – through telephone calls, emails to specific individuals, bringing people together, processing data with them, etc. This renders the work somewhat invisible at times, which makes it hard for people to understand what you're doing. It's not always clear what you are doing or how you are accountable. They can't track it. It is easy to produce reports to demonstrate something but that may not be the most useful approach. This makes a DE role a bit fuzzier.

Study participant M reports struggling with a similar challenge and reports that many evaluation users may be reluctant to pay for necessary activities such as frequent communication between the evaluator and evaluation users:

> It's also difficult to budget for the close and pretty frequent communication with the client. They often ask why we've budgeted so much for management. It's hard to put it into a line item without it sounding out like unnecessary overhead.

Finally, adapting evaluation plans and activities to reflect the shifting priorities of evaluation users also means high transaction costs, requiring evaluators to drop areas of work in-development and begin the messy, inefficient and often undervalued work of starting something new. Study participant R describes a series of dramatic changes in the evaluation of a network of leadership development programs in the midst of a rapid expansion in members:

> In one quarter, the work on network impact was the thing to do but we changed the next quarter because we decided to run an annual retreat.

Finally, study participants reported that it is sometimes equally challenging to immediately spot or describe the outcomes of good evaluation work. How do evaluators and evaluation users know when evaluative activities have led to better thinking or decisions? Study participant K, for example, raised the following question.

How do you help evaluation users understand the value of what they are getting in evaluation – particularly developmental evaluation? I personally have a hard time telling people what it's going it cost and when they do, they never expect it. There has to be a better way of going about this.

The very nature of these challenges means that developmental evaluators are periodically unsuccessful in convincing evaluation users to agree to cover, or, to accommodate changes that entail additional costs. Study participant I described several situations in which he agreed to "eat the difference". Study participant F questioned whether developmental evaluation was economically possible for an independent consultant whose clients were largely are non-profit organizations with typically modest resources available for evaluation.

(d) working across scales

Some study participants also describe challenges associated with working effectively with different users across different scales of an emergent intervention.

This is particularly difficult in large interventions where there are multiple evaluators working on multiple levels of action. Study participant G, for example, described how his evaluation team for a national network of urban collaborations aimed at reducing poverty is able to support the core staff and volunteers integrate developmental evaluation into their preparation, monitoring and ongoing adaptation of their overall strategy, but does not have the resources to integrate developmental evaluation at the level of thematic strategies (e.g. workforce development) nor specific initiatives (e.g. an effort to change a policy): there are simply too many initiatives (i.e. over 130), with too many diverse stakeholders, spread across too many sites. This means that the national evaluation team has only a limited knowledge of the ground level activities and effect of the poverty reduction initiatives.

Study participants F, M, R described cross scale initiatives where different evaluation teams carried out their work independently of each other.

Sometimes these evaluative processes unfolded independent of each other. For instance, study participant D described how in spite of their initial intentions and considerable effort, she was unable to link the evaluative work of the operational group managing a trial set of

innovation-inducing exhibits and developing an innovation framework with the leadership group concerned about innovation across the museum overall.

> When we began to employ DE more systematically and explicitly, it was at the operational level with the exhibit group. That was fabulous. However, I felt that we inadvertently did not tie together the two levels. Had we continued to involve the project leadership throughout this part of the initiative, the work may have had a greater effect on them and the broader organization. It may have led to more time, funding and perhaps energy for continued DE.

Study participant D described a variety of reasons for weaker-than-anticipated linking across scales. Some of them were because the conditions for developmental evaluation were weak: the evaluator was relatively new to developmental evaluation and was eager to experiment with the approach before she began work with the leadership group; the group sponsoring the developmental evaluation was the evaluation team – not the leadership group – which made it difficult to get buy-in; the leadership team was working with short-time-lines to get the exhibit space set up and operational. The other reason relates to the nature of developmental evaluation: the evaluator and operational group became consumed with developing, testing and refining the innovation framework and exhibits, leaving little time for engaging other leaders.

5.8 Evaluator Capacity

5.8.1 Attributes

What are the characteristics of a developmental evaluator? The majority of the 18 study participants identified a variety of characteristics organized broadly into three themes: hard skills and knowledge, soft skills and knowledge and personal attributes, (See Table 5-7). These themes are explored in more detail in the following sections.

(a) hard skills & knowledge

One-half of the study participants felt that evaluators required a strong background in evaluation to be effective in developmental evaluation. Three out of the four study participants who did not have formal training and education in evaluation, for instance, reported that their

evaluation work would be stronger had they had formal training in research methods and evaluation.

Table 5-7
Attributes of Effective Developmental Evaluators

AREA	PARTICIPANTS	N=14[33]
Hard Skills & Knowledge		
Strong Research Methods	A, B, C, E, G, H, L, M	57%
Understand Evaluation	C, E, G, J, P, R	43%
Domain Expertise	C, G, M	21%
Experience in Innovation/complexity	P, K, I	21%
Able to Make Data Accessible	B, F, K	21%
Project Management Skills	R	7%
Soft Skills & Knowledge		
Facilitation	A, E, Q, R, G, H, I, R	57%
Listening	D, K, J, L, M	36%
Communication	D, G, I, J, R	36%
Interpersonal	A, G, M, J	29%
Speak Truth to Power	F, G, I, F	29%
Pattern Recognition	A, F, I, J	29%
Ability to See Forest & Trees	B, D, F, G	29%
Analysis/Synthesis	G, R	14%
Conflict Resolution	R	7%
Team Work	J	7%
Personal Attributes		
Comfort with Ambiguity & Paradox	D, E, I, M, J, R, G	50%
Humility & Confidence	A, B, M, Q	29%
Flexible/Adaptive/Quick/Creative	D, K, G, J	29%
Belief in Evaluation	B, G, J	21%
Experience with Innovation	K, P	14%
Commitment to Issue	R, G	14%
Outcomes Orientation	I	7%
Practical	B	7%
Ability to Motivate	A	7%

Four participants added that the evaluator's command of research method should be strong enough that they are able to design, implement and adapt methods quickly in order to reflect the emerging and sometime fast pace of developmental processes. Three participants stressed the need for evaluators to be able to make the data 'real', 'meaningful' and 'practical'.

---

[33] Four of the 18 participants in the study did not answer this set of questions.

Study participants did not agree on whether having domain knowledge of the intervention in question was important. While four study participants felt that domain knowledge as an important attribute, others were not as sure. Study participant C, J, K and M, for example, argued that as long as the evaluation users understand the domain in which they work, the evaluator's job is to encourage them to tap into that knowledge, rather than supply it him/herself. Study participant D and I, on the other hand, felt that domain knowledge was helpful, but not critical.

(b) soft skills & knowledge

Study participants most frequently identified facilitation as the most important soft skill required by an evaluator. This was because of the need of evaluation users – rather than the evaluator – to assume the lead role in thinking and acting evaluative, gathering and making sense of information and using that information to make decisions. As study participant Q noted:

> You need to have extremely strong group facilitation and adult education skills
> so you can help with decision-making, quickly put together group processes that
> are efficient and effective, communicate effectively, resolve conflict, etc.

Study participants identified a range of skills related to facilitation: this includes good listening skills; strong oral and written communication; good interpersonal skills, including an ability to work with a diverse range of people ranging from high level decision-makers, front line workers and program beneficiaries; an ability to work in teams; comfortable with speaking truth to power, and basic conflict resolution skills.

Several study participants also stressed the need for the evaluator to have a natural ability to recognize patterns and to observe when there was convergence and divergence of data. Because developmental evaluations often unfolded at multiple levels – particularly in the case of interventions aimed at changing systems at the community, organizational and program level – three evaluators felt it was important for the evaluator to be able to think and act across scales and work with the "big picture and little picture" and see the "forest and the trees".

(c) personal attributes

Hard and soft skills are one thing; innate qualities are another. Study participants also identified a number of personal attributes they felt an evaluator needed in order to be effective in developmental situations.

Study participants K felt that developmental evaluation required someone with "a different innate temperament and talent than required in other types of evaluation". Approximately one-half of interviewees reported that this included being comfortable with the ambiguity, uncertainty and paradoxes that typically accompany developmental situations.

Being an effective evaluator may also require a balance between self-confidence and humility. Study participants B and Q argued that while evaluators they must have sufficient authority and credibility to inspire confidence in evaluation users, they also need to let go of being the expert and work instead in the "client's space". "It's not about you, it's about them" remarked study participant B, an orientation which study participant Q termed "humility" and a practice study participant L called "servant leadership".

5.8.2 Emerging Questions & Issues

(a) individuals or teams

Given the ideal characteristics of a developmental evaluator and the broad range of unique circumstances in which they find themselves, many study participants discussed uncertainty about the extent to which a single evaluator is able to fulfill the role of developmental evaluator on their own.

The majority of study participants appeared to feel that developmental evaluators are more effective operating as part of a team or a network. Study participant Q, for instance, notes:

> You need a very broad spectrum of networks, resources and knowledge on which to draw so you can bring in the right people, the right information (in and out) in a timely way. One person cannot be an expert at everything.

The ability to work in teams may be particularly important in large-scale evaluations. Study participants C, E, J and M felt that teamwork was unavoidable for larger assignments which involved multiple layers of evaluation and required a wide variety of skills and perspectives.

Study participants I, K, Q and R noted that evaluators should establish networks of colleagues with diverse skills and areas of domain knowledge that they can quickly call upon as needed for smaller assignments. However, building and sustaining such networks may be easier for evaluators working in larger organizations than those working independently as self-employed consultants. Study participant Q, an employee of an organization that makes extensive use of evaluation, describes their ongoing commitment to connecting with others in the field:

> I am constantly meeting with people in different outcome areas in which [our organization] is working, so I can bring information and relationships into the picture relatively easily and quickly. I might have to hire someone to do some analysis, have a speaker, or outside resource, but I can do this because I have these connections.

Study participants I, R and G, self-employed contractors, on the other hand, described how they felt they had relatively few resources and structured opportunities to develop and maintain communities of practice and networks of colleagues, particularly with colleagues with whom they might eventually compete for evaluation contracts.

(b) can anyone be a good developmental evaluator?

The list of attributes for an effective developmental evaluator is large. Study participant Q and K argued "you need everything a traditional evaluator needs plus more", a sentiment echoed by study participant C:

> I think it's highly sophisticated form of social inquiry. For it to be successful it requires a high level of skills and expertise, a real awareness of what you do/do not know, can/cannot do, a degree of credibility and legitimacy. It might sound snobby, but it's on the tougher end of the competency spectrum.

Study participant I went a step further and argued that there was likely only a small pool of people with the necessary mix of attributes required to be effective developmental evaluator:

Very few people can be very good at this. This is not surprising: there are very
few people that are really good at anything. Can anyone be a doctor? No. And
not every doctor is an excellent doctor. You can say the same for economists or
teachers.

Study participant R agreed, arguing that the uncertain and emergent nature of
developmental evaluation meant that only a portion of conventional evaluators were suited to the
work: "I have to say that this type of work is not for every evaluator. Many would be driven
crazy. It's too messy".

(c) developing developmental evaluators

Many study participants wondered aloud about the extent to which developmental
evaluators could be developed through education, training and mentoring.

Some study participants felt that people could be educated and trained in developmental
evaluation, but that ultimately learning-by-doing and first-hand experience was the most
effective way to develop the capacity to be effective. Study participant C, for example, was not
convinced that "someone just out of school" had enough life experience to be an effective.
Study participant I felt that someone older was often more comfortable in handling tension-filled
situations, negotiating roles, etc. with the dynamic nature of developmental situations. Study
participants B and I argued that experienced evaluators were apt to be more effective because
they had greater credibility in the eyes of evaluation users.

On the other end of the spectrum, some study participants felt it was possible to
have too much experience. Participant K notes, "On one level, when you are younger,
you don't have as much experience, confidence, etc. On the other hand, you are not
steeped in your ways and may be more adaptable."

Study participant E felt that the truth was somewhere in the middle younger, less
experienced yet more adaptable and older, more experienced and rigid:

I am coach to several local evaluators in the region. A lot of them have been in
the field for 20 years. I find that evaluators with 8-15 years of experience are most
comfortable with developmental evaluation and its adaptive approach. Those
with more experience, that have their own tried and true methods, tend to worry

about methodological rigor and don't like changing instruments mid-stream because it interferes with that rigour. On the other hand, my intuition is that younger evaluators may be more flexible, but don't have the experience, knowledge and skills with past projects that allow them and willingness to be flexible, ability to draw on past projects.

Several participants felt that while the hard and soft skills of developmental evaluation could be trained, the personal the attributes required for effective developmental evaluation were innate and therefore not teachable. Study participant F, for example, asked, "Can you teach someone to be an entrepreneur?" She continued the analogy and proposed that while it was possible to help someone develop the skill of book-keeping or marketing, it is not clear that it was possible to teach them to be emotionally prepared to take risk, be self-motivated, and see opportunities where others see none.

5.9 Conditions for Developmental Evaluation

Not all situations are ripe for developmental evaluation. The following two sections describe the conditions interviewees feel are required for effective developmental evaluation, some practices employed by them when conditions are not ideal and emerging questions.

Table 5-8
Conditions for Effective Development

AREA	PARTICIPANTS	N=14[34]
Ability to Work Adaptively		
Urge for change.	Q, I	14%
Comfort with innovation/complexity	K, F, A, M	29%
Authority to make decisions	A, B, I, J	29%
Adequate resources	I	7%
Flexible and Adaptive	A, D, I, M	29%
Tolerance for Risk/Failure	Q, O	14%
Adequate time	A, F, R	21%
Evaluative Thinking & Practice		
Positive Experience with Evaluation	O	7%
Commitment to evaluative thinking & data	A, D, E, I, O, R	35%
Understand Developmental Evaluation	D, E, R	21%
Committed & Stable Leadership	A, E, H, J, R	35%
Engagement of entire development team	K, M, O	21%
Adequate Time & Resources	Q, E, R, H, M, I	42%
Trust of Evaluators	A, I, C, E.	29%

---

[34] Only 14 of 18 study participants provided feedback to these questions.

5.9.1 Stakeholders Ability to Work Adaptively

Approximately one-half of interviewees felt that the success of developmental evaluation depended on whether the users of the evaluation were working with conditions that allowed them to "develop" something. Speaking from a variety of experiences, they described the following characteristics:

- An urge to "change" a condition. The motivation to make a change is important to work through the messy, uncertain, iterative and often failed process of developing and adapting a solution.

- An understanding of the dynamics of development. Development processes unfold more naturally and easier when people understand the inevitability of uncertainty, iteration, failure and adaptation.

- Authority to make decisions. Developers have the authority to make decisions to change any part of the development (e.g. problem definition, theory of change, design) based on new learning's, stakeholders and changing contexts.

- Adequate Resources. Developers have adequate resources to try new things, explore multiple paths, make mistakes and mid-course adjustments.

- Flexibility & Adaptive. Developers are ready, willing and able to continually adapt in the face of new learning's.

- Tolerance for Risk & Failure. Developers are comfortable with the prospect that they will experience any false starts, dead ends and failures en route to developing something with a chance of working.

- Adequate Time. Developers have sufficient time to allow something to emerge and develop rather than working with artificial or short term timelines.

Approximately one-half of interviewees felt that the success of developmental evaluation depended on whether the users of the evaluation were working with conditions that allowed them to develop something

5.9.2 Evaluative Thinking and Practice

While the conditions for good development overlap with the conditions for effective developmental evaluation, interviewees described a number of features related to evaluative thinking and practice that they felt were helpful:

- Positive experiences with evaluation. People with negative experiences with evaluations and evaluators efforts to enhance their evaluative thinking capacities may be suspicious about DE. Those with positive experiences will be more open to it.

- Commitment to evaluative thinking. People who are prone to reflect on and test their beliefs, ideas and theories based on data, critical thinking are more likely to demonstrate these in ambiguous and fast moving developmental contexts.

- Knowledge and understanding of developmental evaluation. People that know about and understand developmental evaluation are more likely to use it productively than people without prior knowledge or misconceptions of the concept and its application.

- Committed and credible champion(s) for DE. Having someone to advocate for integrating developmental evaluation into the day-to-day work of the people developing the intervention; "leadership adhesion".

- Adequate time and resources. It takes time, energy and resources to gather and critically reflect on data on the developmental process.

- Trust of evaluators. The effectiveness of developmental evaluation is stronger when there is trust between evaluators and evaluation users.

Two study participants (B, J) reported that while these conditions are typically helpful in any evaluation process, they are extra helpful in developmental situations where the fast moving pace of development makes to interpret the world and make decisions without the benefit of data and critical thinking.

5.9.3 Common Challenges

The conditions for developmental evaluation are ideal but they are rarely fully present in all situations in which developmental evaluation is being considered. Interviewees identified five broad overlapping challenges they face in employing developmental evaluation effectively.

(a) ecologies too rigid for adaptive development

Study participants described the challenge of trying to carry out emergent and adaptive work in rigid ecologies or settings. Study participant B argued that evaluators should pay close attention to the extent to which their clients had the authority to make decisions about the program:

> If we want to help people adapt their approach, we have to realize that they are working in complex organizational settings. They don't always have the degree of control or influence (even the funders) on the program to make much changes and it's very difficult to work under the spotlight in rigid settings.

Rigid ecologies refer to situations in which intervention stakeholders – and those that support them - operate in a culture and set of policies and rules that require an intervention to be rooted in a clear understanding about the cause and effect relationships of the issue they are dressing, reflects an explicit theory of change, relative stable design, predictable contexts, unfold on fixed timelines and generate relatively predictable outcomes. This is not the ecology required for emergent and rapidly changing interventions where the cause and effect relationships underlying problems are often unclear, theories of change implicit and emerging, operating in changing contexts, employing multiple experiments, unfolding with unpredictable timelines, and are as likely to fail as succeed.

(i) a preference for linear approaches

Several study participants reported that their effort to provide evaluation assistance to innovators was made more difficult because they approached "complex issues as if they were simple". Study participant F felt that many people working in complex and innovative contexts struggle because they approach the challenge in a linear fashion, trying to clearly define problems, explore and choose options, and then get into practical issues related to conceiving and elaborating an intervention. The reality, they argued, was that the process was unavoidably messy and iterative with more questions than answers.

> Most of what we do is at the conventional level and we don't have a lot of tolerance for the kind of learning and development that is required to support [adaptive] diagnostics and design.

Study participant K agreed and argued that developmental evaluators often needed to help educate those developing new approaches about the dynamic nature of this process, which she called adaptive management.

> Even people that are natural adapters are not always clear about patterns of adaptive management. This is why you need to bring sensitising concepts like complexity, eco-cycle frameworks and examples of innovation to reinforce how this stuff actually happens".

This lack of understanding can create challenges for developmental evaluation. This includes developmental processes that are highly theoretical and abstract when developers feel that they must fully develop and design something conceptually without "getting their hands dirty" trying stuff out on the ground (Participant I) or asking evaluators to build elaborate and expensive strategies for gathering and analysing data when the intervention – and appropriate evaluation design – are apt to emerge (Participant C).

(ii) an urge to over-specify

A rigid ecology can manifest themselves in a variety of ways. Sometimes it means that external authorities over-design the intervention very early on in the process. Study participant I argues that this creates a dynamic of path dependency which leaves little room for people on the ground to experiment and adapt the initiative as it unfolds[35].

Study participant F described how the early decisions by funders of a national project to support local efforts to mobilize youth in civic affairs made it difficult to explore new ways of approaching the work uncovered by the evaluation:

> Things that were set in motion at the start that made it very hard to shift things later one: kind of like a game of chess, where how you play the game to a point in time determines what kinds of options you have moving forward. So even though [the local organizations] learned things that would help them to move forward in more effective ways, they did not have the capacity to shift their work because some of the things that had already been put into play limited what they could do with the communities. […] You need to have the capacity

---

[35] Definition of path dependency from Google wikipedia: when the decisions one faces for any given circumstance is limited by the decisions one has made in the past, even though past circumstances may no longer be relevant.

and opportunity to be able to integrate and implement the learning – and that doesn't exist in all cases."

Even when stakeholders do not have a design of their own in mind, their urge for certainty can prompt them to require those developing a new approach to produce a relatively well-defined intervention and plan before the actual process of developing that intervention has even begun. Study participant A, for example, described how an international donor organization asked grant recipients for a fairly detailed account of how they intended to strengthen the agricultural sector in Africa:

> The funder asked us to provide them a completed logic model and workplan before we even began our work. We told them that were unable to do that right away, that it would emerge over time as we worked together and experimented. But their funding protocol and accountability measures required that we submit these things before we started. That put us in an awkward position.

She went to describe how the team spent a "disproportionate" amount of time meeting the funder's need for specificity while trying to keep the space open for flexibility on the ground.

(iii) short and fixed timelines

It is difficult to plan, implement and complete the process of development or significant adaptation on a short and fixed schedule. It may also be inappropriate. Study participant Q argued that any interventions "always slice in and out of a larger and longer term development process". And while they may make the work of funders, administrators and evaluators logistically easier, she continued, they do not reflect the nature of development work and can be counterproductive.

Sometimes funders and administers impose the timeline. Study participant A argued that standard practice of time-limited project funding had the mechanical effect of limiting the breadth, depth and duration of experimentation by everyone involved.

> [Our] projects begin to converge mid-way through the work because funding timelines create clear end points. People experiment less –even reflect less – because the project is almost over and they need to wrap things up.

At other times, it is the innovator who establishes artificial and inappropriate timelines. Study participant F summarized her experience working with many clients who profess an eagerness to innovate but only within fixed timelines: "[They] want the kind of results that come from adaptive learning, but they want them on a time schedule that is more consistent with conventional learning".

(b) user disinterest in evaluation

Even when the conditions for organic development exist, many intervention stakeholders are unwilling or able to subject their hunches, beliefs and theories about how to make a difference to a rigorous process of evaluation.

Sometimes it's because prospective evaluation users are far more motivated and informed by beliefs and values than by knowledge or data. Study participant R concludes that "even reflective practice is a stretch for people that are action-focused". Study participant O described the difficulty involved in introducing evaluative processes into what he described as one of Canada's most mission driven charitable organizations:

> [The organization for which I am the evaluator] does not have a reflective bone in its body. They are an amazing organization doing amazing things. But they blindly follow their mission and principles without reflecting much on what that means or how the effects of their day-to-day work play out for people. I once asked if we could evaluate how well a team meeting had gone and they looked at me like I was crazy.

Yet other times, prospective evaluation users are interested in a more rigorous evaluation process, but not interested enough to invest the time and energy required do it effectively. Study participant F described how a key project manager was reluctant to approve the resources required to get feedback on the early effects of the group's work. "He seems to get DE but then asks questions about whether some of the DE work I want to do (e.g. key informant interviews) is worth the effort".

Finally, sometimes the reluctance of prospective users of developmental evaluation is rooted in a previous bad experience with formal evaluation. Several participant described how clients complained about participating in evaluation processes that were "overly judgemental",

and/or offering information that was "unhelpful", "poorly communicated" and "too late to make a difference".

(c) turnover in evaluation users

Many study participants (A, E, I, K, Q, R) described instances where the departure of key evaluation users – and/or the arrival of new users – meant that the development process and/or evaluation activities were "put on hold" or "had to start all over again". Study participant R described a situation how early on in a project, her primary accountability was to the senior manager who brought her into fast growing project on leadership development for young women. This changed with the arrival of several new managers hired to deal with the explosive growth of the program, requiring the establishment of entirely new relationships.

Study participant O argued that turnover in evaluation users may present challenges for developmental evaluation but that they are a normal part of developmental contexts and therefore should be fully expected in ever-changing interventions.

5.9.4 Dealing with imperfect conditions

For the majority of interviewees, working with less-than-ideal conditions for developmental evaluation appears to be the norm rather than the exception. Study participants described four options in such situations.

(a) don't start

When the conditions for effective developmental evaluation are sufficiently weak, evaluators and evaluation users have the option of deciding not to introduce a formal evaluation component to their work. Study participant B with over twenty years of experience in the field described that she and a prospective client have a "fit" in a relatively short period of time:

> I look for where the group is at and how I can help. I am only interested if I can help. If I can't help, we don't work together.

In other cases, the process of determining whether the conditions for proceeding exist takes a much long time and may result in a decision not to proceed. Study participant I described

his interactions with a large human service organization that approached him to assist them with use DE in the development of a new venture:

> They were well down the road on conceptualizing and developing [a large new project to assist homeless persons]. They wanted to [employ DE] and we spent months doing things but spun our wheels trying to find an area to develop. We would find something that was technically suitable, but they would say it's not super important. We'd then find something else and it would repeat. There was nothing we could find that they had an interest and appetite to explore using DE.

The study participant went on to describe how their efforts to create a new program "petered out" and he and the client eventually agreed to "try again" at a later date.

(b) improve conditions as the work unfolds

Another approach to dealing with less than ideal conditions for developmental evaluation is to carry out evaluation activities in a way that strengthens those conditions. This option was described by a person for whom building evaluation and learning capability of their clients was a core objective of all their work. She described a process where the strength of the conditions ebbed and flowed continuously:

> Conditions change in most of our work. And certainly the longer the project, the more likely it is that conditions will change (e.g. projects for over five years). Even if the conditions change for the worse, I don't think evaluation work – even DE work – is no longer appropriate. It simply means there are new obstacles to acknowledge and address. You may lose the conditions for developmental work for a short period of time before you build them up again.

While the strategy used to improve the capacity of the client depended on the situation, she went on, her organization typically facilitated this by transferring more and more evaluation activities from their evaluator to the organizations or networks over time, with the evaluation firm acting as "coach". If and when the conditions temporarily worsened – e.g. loss of internal staff persons with key role in the evaluation - the evaluation firm would temporarily re-assume responsibilities and roles until such time as they could be transferred back once again.

145

(c) cocoon the process

When conditions for emergent development and developmental evaluation activities are weak, evaluators and evaluation users can attempt to cocoon the process. This option was described by a study participant who had extensive experience working with international development organizations, both in the field as well as with central agencies, and who characterized doing developmental evaluation in weak conditions this way:

> The conditions [for developmental evaluation] do not always exist in international work where the emphasis is on the big plan, controlled inputs and outputs, knowing exactly what should – and can – be done, predictable results, etc. This is the mainstream paradigm in the development industry and it's reinforced by results-based management. We need to step out of that approach and say we don't know everything we need to know to make this approach work – and we may never know.

The study participant used the term cocooning to describe a project in which a "champion for organic development created and kept open a space for developmental evaluation" by ensuring that the requirements of the international donors were met without undermining the work in the field. S/he illustrated this process first by describing how she and her team an dealt with a funder's requirement to provide a project theory of change, evaluation design and set of outcome measurements, and budget required by their Results-Based-Management framework early on in what was an emergent, grass roots approach to expanding the use of agricultural technologies in a region:

> [Our funder] pushed us … they wanted more clarity on what we were going to do and how. We responded that we didn't know that yet, where the project would go, but we did know some basic things: staffing, operational costs, a funding mechanism. We described these things and our basic process as clear as possible. They relaxed a bit. This seemed to work. I don't know if you can get away with that now with the stronger push on accountability and results but it worked out when we did it.

The same study participant continued to describe how the practice of cocooning may require evaluators and evaluation users to carry out concurrent evaluation processes, each designed for different users, answering their preferred questions, employing their preferred approaches, and reporting in their preferred manner.

> We did accommodate and make a shift that allowed us to develop and use methodologies where could extract information to give to [our international donor] for their Results-Based Management template. For own purposes, we use Most-Significant-Change methodologies and reflective reports which generate information but help use with decision-making, but not things need to report. In the Most-Significant-Change example, our [local] partners and participants each reviewed each other stories as part of the process. It was fascinating for us but did not easily fit into RBM. So, we have two evaluative processes operating alongside each other that sometimes overlapped.

The practice of cocooning can be strengthened with the active involvement of the administrators and funders of the initiative under development. Study participant C, for instance, describes how "skilled bureaucrats can work the system for you". This might include a strategy of "nods and winks" where they sometimes "quietly recommend" that we "just fill in the form but don't tell me what you do". Study participant B pointed out that funders and authorities who make an extra effort to "navigate the system" and free up space for emergent work may pay a heavy price: "Good funder-grantee work now usually depends on a heroic person who does too much. When that person burns out then you have to find a new one."

(d) discontinue or transition the evaluation

Finally, when a situation is still technically developmental but the conditions for effective developmental evaluation change for the worse and even the best capacity building and/or cocooning strategies are no longer effective, evaluators and evaluation users can choose to discontinue the evaluation activities.

Sometimes the evaluator makes the choice to discontinue the developmental evaluation activities. One study participant described her decision to leave a project when the window for real development and developmental evaluation appeared to close significantly:

> I asked to be taken off [the project] because the conditions changed. When we lost our program coordinator with only six months of funding remaining, the local partner decided not to simply embed some of the [emerging strategies] into their existing programming. I didn't get the sense that this was going to involve much adaptation of their existing programs. So, as the learning requirements seemed to be minimal and we were no longer in a development phase, I determined that there was little need for a DE.

Sometimes it is the evaluation users who decide to wrap up the evaluation work. One interviewee described an occasion in which a client terminated the evaluation contract with the evaluator because she had not stuck to the original work plan despite the fact that in both cases the evaluation users and project had changed a great deal.

 Finally, *s*ometimes developmental evaluation work winds down on its own without a conscious choice by either the evaluator evaluation user. Study participant K describes how evaluation users suddenly shifted their attention elsewhere:

> I think the proportion of energy focused on the DE piece of it lessened or decreased. Someone suddenly comes in and asks you – the evaluator -- to answer some more precise questions, generally about the extent to which something is working. They only have so much money, so much time and/or need to make some decisions. It's natural.

The shift can be slow and imperceptible. Study participant D described how the developmental evaluation activities for the ever-evolving museum exhibits tapered off after most of the original staff team moved to another organization and the museum did not renew the budget for evaluation. Study participant I described how the effort "withered away slowly and almost imperceptibly" as the initial probes and experiments to develop a new approach to environmental education failed to surface possible strategies the organization felt were worth pursuing.

5. 10 Conclusion

The first-hand experience of the 18 early adopters of developmental evaluation interviewed in this study largely reflects the overall concept and general features of the idea and approach established by Patton. Study participants do, however, have insights and questions about the theory and practice of developmental evaluation that reveal

The majority of study participants feel that developmental evaluation does represent a distinct evaluation niche. This includes focusing on the distinct purpose of assisting intervention designers "develop" an intervention as well as the various contexts in which this type of approach is required (e.g. pre-formative, cross-scale interventions, etc.). While some participants feel the practice of developmental evaluation is similar to other disciplines (e.g.   planning and

organizational development) and methods of inquiry (e.g. social learning and participatory evaluation) they still describe their work as evaluation

Study participants appear to manifest the ideal roles and relationships of developmental evaluator envisioned by Patton: i.e. asking evaluative questions and seeking out just-in-time feedback as a team member developing an intervention. At the same time, some report confusion and practical struggles in this area. The most significant challenge is understanding and navigating issues related to accountability specifically, to whom the developmental evaluator accountable and for what is s/he accountable. Several participants pointed out that Patton does not explore these issues in much depth.

Study participants' approach to methods in developmental evaluation matches the orientation prescribed by Patton. They describe using a variety of sensitizing concepts to assess the degree to which a situation may be appropriate for developmental evaluation. They also employed a contingency approach to methods and an adaptive strategy for design which allowed them to shape the evaluation design to different situations and to adjust to reflect changes in the interventions, their users and shifts in contexts. For a variety of reasons, several participants expressed a bias for qualitative methods.

Table 5-9

Summary of Interview Findings

AREA	ORIGINAL CONCEPTION	INTERVIEWEE REFLECTIONS & EXPERIENCE	
		ALIGNMENT	**EMERGING ISSUES**
PURPOSE			
Distinctions	Emphasis on developing – rather than improving or judging – an intervention.	Emphasis on developing – rather than improving or judging – an intervention.	Some confusion and new insight about the distinction between formative and developmental processes and evaluation.
Niches	Pre-formative, constant adaptation, cross-scale complexity, crises situation, replicating intervention.	Experience with pre-formative, constant adaptation and cross-scale complexity; no experience with crises or replication.	Suggestion that developmental evaluation can be a "probe" to encourage strategic thinking.
Other Disciplines	Similar to organizational development.	Similar to organizational development and planning.	Minor confusion about the distinction between strategic learning and development evaluation.
EVALUATOR ROLES			
Roles	Framing issues, testing quick iterations, tracking developments; challenge to keep a result focus.	Framing issues, testing quick iterations, tracking developments; challenge to keep a result focus.	No discussion about challenge of keeping results focused; concern about getting involved in the actual development and questions about the role of evaluator in judgment about developments.
Relationships	External or internal member of the development team.	External or internal member of the development team.	The extra importance of trustful relationships.
Accountability	Accountable to development team to facilitate evaluative thinking; aware that power of funders and decision-makers.	Accountable to development team to facilitate evaluative thinking; real difficulty in dealing with power of funders and decision-makers.	Competing paradigms of accountability; challenged by high turnover of stakeholders; questions about what constitutes a "deliverable".
METHODOLOGY & DESIGN			
Situational Recognition	Use of sensitizing concepts, such as complexity and panarchy.	Use of sensitizing concepts, such as complexity and panarchy.	Mix of structured and unstructured processes.
Methods	Emphasis on good questions, participatory processes, contingency-based, adaptive; challenge of dealing with large volume of data.	Emphasis on good questions, participatory processes, contingency-based,, adaptive; challenge of dealing with large volume of data.	Slight preference for qualitative methods; questions about burden of proof and reliability and validity of data.
Design & Budget	Adaptive and flexible design. Few direct references to issues of budgeting.	Adaptive and flexible design; higher-than-normal transactions costs and accounting for "invisible" work; challenges with turn-over and diverse stakeholders.	Four different approaches to adaptive design and budgeting.
EVALUATOR CAPACITY			
Skills, Knowledge and Attributes	Same competencies required by traditional evaluation; comfort with ambiguity and uncertainty; domain expertise helpful; challenge of remaining credible.	Same competencies required by traditional evaluation; comfort with ambiguity and uncertainty; different perspectives on the importance of domain expertise.	Suggestion that developmental evaluators stronger as part of a team; several feel that not all evaluators can be effective and/or trained to do effective developmental evaluation.
CONDITIONS FOR DEVELOPMENTAL EVALUATION			
Evaluation Users Able to Work Adaptively	Few direct references to issues of adaptive leadership and management but described consistently through publications.	Some evaluation users – or their senior decision-makers or funders - prefer linear, over-specified development with short timelines creating a rigid ecology for adaptive development.	Four approaches to working in imperfect conditions.
Evaluation Users with Evaluative Capacity	Few direct references to issues of evaluative capacity of evaluation users but a general theme in utilization-focused evaluation.	Some unwilling to submit hunches and ideas to "reality testing"; challenge with high turnover	

The emergent nature of developmental situations appears to amplify the traditional constraints typical in evaluation practice. This includes the comparatively high rate of turn-over of evaluation users, the often messy politics and perspectives of diverse stakeholders working with complex issues and evolving interventions , and the need to continually adapt the design and budget for the evaluation to reflect shift in the intervention and contexts. Several participants shared their challenges and questions about what constituted practical "burden of proof" as well as "reliable and valid data" in developmental situations.

Participants' reflections on the broad capacities that evaluators require to be effective in developmental situations roughly fit those described by Patton. This includes the basic competencies required by any evaluator, complemented by a comfort with ambiguity and uncertainty and some domain expertise related to the topic of evaluation. Several of them go farther than Patton, however, as they question whether everyone has the inherent capacity to be a good developmental evaluator and express uncertainty about the extent to which these competencies can be developed through training and education alone. Given this "higher-than-normal bar of ability" required in developmental evaluation, several participants recommended that evaluators be prepared to employ participatory processes that tap into the insight and evaluative capacities of evaluation users, (b) works in teams if possible and (c) cultivate "networks" of evaluators they can call on an as-needed basis.

Finally, the experience of early adopters is that the conditions for effective developmental evaluation vary widely. While study participants shared examples of "social innovators" eager for data and critical thinking as they navigate the messy, ambiguous, non-linear and uncertain process of developing or adapting an intervention, they also described instances of evaluation users who were not keen to test their hunches, beliefs and interventions with evaluative processes. Moreover, they describe how the ability of many innovators to work adaptively is constrained by the "rigid ecologies" in which they work. This includes organizational cultures, policies and practices which (a) encourage a linear (rather than iterative) process of exploration and design, (b) fixed and short (rather than open and long) time lines, and (c) greater (rather than less) specificity about what is meant to be developed.

Study participants demonstrated their ability to adapt when they shared a variety of approaches and examples of how developmental evaluators and evaluation users can overcome imperfect conditions. This includes the intriguing concept of "cocooning", a developmental process in which social innovators seek to work adaptively while concurrently actively trying to minimize the pressure to work otherwise from administrators, decision-makers and their own team members.

Chapter Six: Conclusions & Recommendations

6.0 Introduction

The chapter returns to questions that guided this study, offers a set of conclusions about the conception and experience of developmental evaluation based on the study findings, proposes recommendations that are focused on strengthening the theory and practice of developmental evaluation, and closes with some final remarks about the place of developmental evaluation in the larger field of evaluation.

6.1 The Problem Statement, Objectives and Approach

The problem statement of this thesis is composed of three questions: What is the current theory of developmental evaluation and how does it differ from other evaluation approaches? What are the experiences and reflections of 'early adopters' of developmental evaluation? How can the theory and practice of developmental evaluation be strengthened based on this experience?

The research methodology used to address these questions consisted of a review of the literature on developmental evaluation as well as structured, open-ended interviews with eighteen people that have experimented with developmental evaluation.

6.2 Study Conclusions

The conclusions of this study are presented in three sections: first, general conclusions about theory of developmental evaluation; next, conclusions about the practice of developmental evaluation; and finally, conclusions about the ecology for effective developmental evaluation.

6.2.1 The Theory of Developmental Evaluation

(a) conceptual clarity

The concept of developmental evaluation appears to be relatively clear from the perspectives of study participants.

The majority of interviewees reported that they understood many of the major characteristics of developmental evaluation. This included (a) the distinction between more traditional forms of formative and summative evaluation, (b) the contexts in which developmental evaluation was appropriate (e.g. program replication, fast moving contexts, etc.), (c) the broad role of the evaluator, and (d) the contingency-based and adaptive approach to methods.

There is, however, some confusion on a number of characteristics.

There is some modest uncertainty about how developmental evaluation is different from other types of evaluation, methodologies and related disciplines. The includes: the distinction between formative and developmental evaluation; the distinctions between evaluation and research methodologies and functions that may be employed in developmental evaluation or influence its approach (e.g. action-research or appreciative inquiry); and, finally, the distinction between developmental evaluation and other disciplines with which its practices overlap and/or where the approach might be employed, such as planning and organizational development.

Second, there is some confusion and different opinions about the appropriate roles of and boundaries for the developmental evaluator. While Patton is clear that the developmental evaluator acts as team member who brings evaluative thinking and practices to bear in developmental situations, several study participants expressed uncertainty about whether an evaluator should judge the merit or worth of emerging developments or limit their role to facilitating the evaluation users to judge the merit of the developments. There was a difference of opinion among some participants about when the activities of the evaluator spilled over into the boundaries of simple facilitation and organizational development.

Third, there is confusion is about the accountability of the developmental evaluator. Over one-half the interviewees reported they were not clear about to whom the developmental evaluator is accountable (e.g. the funder, the innovator, the person who signed their contract) and for what they are accountable (e.g. a deliverable, responsiveness, progress of the intervention).

(b) critiques

While study participants raised questions or concerns about the concepts and application of developmental evaluation, they did not offer any fundamental critiques of the approach. This may be a mechanical effect of the study: only participants who had already employed developmental evaluation were asked to participate in the study, which suggests that they were somewhat pre-disposed to participating in developmental situations and employing developmental evaluation.

The exception to this general pattern are the two interviewees who reported that they felt that the practical distinction between formative and developmental evaluation, and their respective focus on improvement and development, was a minor issue or even irrelevant. While they reported that a more fully articulated evaluative approach to developmental situations prepared by a respected evaluation expert added credibility to the work, they concluded that implications for their work on the ground were modest. This feedback confirms the unofficial feedback Patton (1998) has received on the topic and the opinion shared by evaluators who operate with a fairly narrow definition of evaluation (e.g. the traditional formative and summative distinction), a reality discussed in chapter three.

6.2.2 The Practice of Developmental Evaluation

(a) practice extends across many – but not all – the niches

Most of the participants in this study reported working in situations which required a different approach than more conventional forms of summative and formative evaluation. These include: (a) a pre-formative situation in which group was just beginning to develop a theory of change and intervention to address a challenge or problem, (b) when a group does not expect or want to stabilize an intervention but instead to constantly adapt to changes in the context in

which they operate (e.g. demographic shifts in clients for a program), and finally, (c) situations of high complexity where cause and effect relationships were unclear and often emergent.

Study participants did not have any experience with using developmental evaluation in situations requiring (d) the replication of an intervention from one context to another nor (e) using evaluative feedback in a crises situation.

One study participant mused that there might be a sixth niche, that is, using developmental evaluation as a probe to stimulate evaluative thinking in general. While using this appears to be a productive use of evaluation, it does not appear to align with intent of developmental evaluation which is to assist a group "develop" a new intervention.

(b) practice is diverse and emergent

The practice of developmental evaluation is still evolving and diverse. Study participants shared a wide variety of experiences and reflections to all the questions posed in this study. This variation is expected because (a) the concept of developmental evaluation itself is still developing and (b) the practice is meant to manifest itself differently in each unique instance in which it is applied.

While the practice may be emergent, some patterns are apparent. The early adopters interviewed shared relatively similar observations on the conditions required for effective developmental evaluation, the processes for assessing the suitability of DE, and features of methodologies for gathering and making sense of data and their observations on the basic skills required by an evaluator.

(c) traditional evaluation challenges are amplified

Developmental evaluation appears to involve the same challenges that evaluators experience in traditional evaluation. Regardless of whether they were experienced or inexperienced, study participants displayed different preferences for research paradigms, struggled to develop robust evaluation designs in the face of practical constraints related to budgeting, were not always clear about their appropriate roles, relationships and accountabilities,

and they were uncertain about the extent to which the evaluation activities and findings were used by intended users.

In the case of developmental contexts explored in this study, however, these struggles seemed amplified in at least two inter-related ways.

First, evaluators and evaluation users who are serious about evaluative feedback in developmental contexts are even more sensitive to question of what constitutes "good data" and an appropriate "burden of proof" for drawing conclusions and making judgements. This is a challenge identified by others working in developmental contexts. In his books exploring the process of innovation in the business sector, for example, Martin (2007, 2009) argues that the burden of proof for valid and reliable feedback on the dynamics and effects of emerging interventions should be necessarily much lighter earlier on in the innovation process when possibility-abductive thinking is employed than in the later stages when proof-based, analytical thinking is required. Patton touches broadly on the same themes in his earlier work on social innovation (Westley et al. 2006) but does not explore deeply in his new book on the topic (Patton 2010).

Next, the emergent nature of developmental contexts makes the tasks of designing and implementing robust evaluation designs even more difficult than in more conventional formative, summative, monitoring and accountability evaluations where the intervention is relatively stable and fixed. This is because the evolving and unpredictable conception and design of the intervention, coupled with the greater likelihood of turnover of evaluation users, means that an initial set of evaluation question and design may need periodic – even constant – adjustments.

The unique nature of these practical challenges raises important questions for understanding what should be expected from developmental evaluation. What is the appropriate burden of proof in different developmental contexts? Does developmental evaluation require more than a conventional amount of resources to account for constant adaptation? Or does it simply mean that evaluators and evaluation users should be open to the possibility that adaptive developmental evaluation designs are not always able to yield the same quality of data as more fixed evaluation design?

(d) specialized competencies required

Developmental evaluation appears to require evaluators to have the competencies they need for other evaluation approaches as well as some additional personal attributes and soft skills. These "extras" include a comfort with ambiguity, an ability to deal with paradoxes and tension, and very strong facilitation skills. It may also be helpful to have prior experience and exposure to the messy nature of developmental processes.

(e) an interest in strengthening practice

Many study participants are interested in expanding their capacity to carry out developmental evaluation.   The includes: (a) aids and processes that can be used to assess the suitability of employing developmental evaluation; (b) approaches to designing and budgeting developmental evaluation efforts; (c) research and decision making methodologies that may be suitable to developmental situations; and finally, (d) options for communicating developmental evaluation results to primary and secondary users. Interviewees expressed an interest in receiving this information in the form of tool kits, case studies and networking and exchanges with peers.

6.2.3 The Ecology for Developmental Evaluation

(a) the dynamics and conditions of development

One of the unintended results of this study was uncovering participants' insights into the dynamic, emergent and unpredictable nature of developmental situations. This includes the process of creating a new intervention, restructuring interventions to reflect new learning's, adopting approaches to respond to shifts in the context in which intervention stakeholders operates, and intervening to change complex adaptive systems.

Study participants reported that they often work with stakeholders operating within institutional and cultural ecologies where the understanding of the dynamics of innovation and complexity are weak or uneven, there is strong preference for fixed plans, designs and budgets, and a low tolerance for experimentation and risk. This limits the scope and timeline of the developmental process and apt to yield fairly conventional interventions.

Their experiences reflect the sentiment of Canadian evaluators who report that current evaluation practices tend to be "overly centered on performance measures and accountability" (Gauthier et al. 2009, p. 30; Sridharan et al. 2011).

(b) conditions for effective developmental evaluation

There appear to be at least three broad categories of conditions that influence the degree to which developmental evaluation will be effective: (a) the capacity, time and willingness of evaluation users to think critically and gather and use evaluative data, (b) the degree to which the evaluator has the skills, domain knowledge and personal attributes "fit" the context and content of the evaluation; and, (c) the degree of trust, respect and perceived credibility between the evaluator(s) and evaluation user(s). These conditions vary from good to poor, and are apt to co-evolve as the evaluation unfolds. These findings appear to broadly align with the general literature on conditions for effective evaluation for other purposes (e.g. formative, accountability) and Patton's own understanding of the factors affecting use.

6.3 Recommendations

The recommendations presented below are divided into two sections and are aimed at improving the theory and practice of developmental evaluation.

6.3.1 Strengthening the Theory

The recommendations described below are designed to assist theorists and practitioners interested in strengthening the theory of developmental evaluation.

(a) clarify distinctions

The theory of developmental evaluation could be strengthened by further clarifying how, where and how it is similar and dissimilar from other evaluation purposes (e.g. formative), methods of inquiry (e.g. action research, social assessment, appreciative inquiry and participatory evaluation, and professional disciplines (e.g. planning, development management), organizational development. This would make it easier for evaluators and evaluation users to be more intentional and effective in using developmental evaluation in their work.

(b) refine or elaborate some key characteristics

The current conception of developmental evaluation could be improved by refining or elaborating on some of the characteristics that study participants reported were either unclear to them or else posed a challenge in their practice. Some of these include: (a) the role of the evaluator in judging the merit or worth of an emerging intervention, (b) roles that evaluators may feel pressure to perform but that go beyond the boundaries of evaluation, and (c) to whom and for what the evaluator is accountable.

(c) introduce new elements

The emerging theory of developmental evaluation can be strengthened by introducing new elements that address gaps surfaced in this study. Some of these include: (a) a review of major tensions in methodology that are endemic – even amplified – in developmental evaluations (e.g. the reliability and validity of data), (b) a discussion of issues related to the burden of proof or making claims in situations when important decisions with sometimes weak and ambiguous data, (c) heuristics for identifying primary and secondary users in developmental situations, (d) a discussion of what might constitute evaluation deliverables, (e) issues and strategies related to evaluating the effectiveness and use of developmental evaluation, and, (f) options for developing, negotiating and adjusting evaluation plans and budgets.

(d) situate developmental evaluation within broader evaluation debates

Advocates of developmental evaluation could strengthen the theoretical base and conception by describing where and how certain features relate to deeper theoretical debates within the broader field of evaluation.   The three areas identified by study participants that are clearly part of longstanding discussions in the evaluation discipline include: (a) the role of the evaluator in judging the merit or worth of an emerging intervention, (b) to whom and for what the evaluator is accountable and, (c) the debates about epistemology and methods between advocates of positivist, constructivists and pragmatists paradigms.

These elaborations would assist evaluators – and evaluation users – to better understand the assumptions, strengths and weaknesses and implications of different perspectives on key

debates. For example, an evaluator or evaluation user who is unclear about the extent to which s/he should judge the merit or worth of emerging intervention would be able to refer to a document that reviews the different perspectives between evaluation theorists such as Rossi, Campbell, Guba and House.

6.3.2 Improving the Practice

(a) educate evaluators and evaluation users

The theory and emerging practice of developmental evaluation is sufficiently well developed that people and organizations interested in educating evaluators and prospective evaluation users about the approach can document and disseminate it more broadly. Based on the findings of this study, I believe it is worth an extra effort to draw out two key themes in further discussions of evaluation.

The first is to highlight (repeatedly if necessary) and better understand the nature of developmental contexts, which Patton describes as "messy, not orderly, emergent and not controlled", and developmental work which is "an iterative process of experimentation, learning and adaptation" (Patton 2008, p. 33). Several study participants report that they come across evaluation users and evaluators who are uncomfortable with (a) dynamic contexts and (b) adaptive processes of leadership, management and design, are less capable of being effective at development and less likely to benefit from developmental evaluation.

The next theme that deserves extra attention concerns gathering and using data in developmental situations. This includes better understanding what is meant by "good research and data" in these contexts. While Patton explored issues related to methods in more depth in his 2010 book than in earlier publications, much of the focus was on "inquiry frameworks" rather than issues related to data and research methods. The experience of many study participants is that the thinking and practice patterns of evaluators and evaluation users are steeped in more traditional forms of evaluation research and that they struggle to manifest with the same robustness in emergent contexts.

The diverse nature of evaluators and users means that communication efforts should be customized to (a) reflect the realities of different users (e.g. policy makers, agency managers, evaluators, funders), (b) working in different domain areas (e.g. education, public health), and (c) across different scales (e.g. for a project within an organization, a local network, a national initiative in different cities). It should include describe developmental evaluation at different levels (e.g. introductory, a small primer, an in-depth review) using different evaluation modalities (e.g. podcasts, webpage, hard copy). Given the complex nature of developmental evaluation, it should be presented described using anecdotes and cases studies to communicate complex characteristics.

(b) provide evaluators with formal training and coaching

The theory and practice of developmental evaluation may be sufficiently well developed that it can be used to prepare formal training and education curriculum for evaluators.

The natural mechanisms for such training include universities with evaluation courses and formal evaluation programs, professional evaluation associations (e.g. Canadian Evaluation Society, American Evaluation Association, and Australasian Evaluation Society) as well as professional associations that employ evaluation (e.g. Canadian Institute of Planners, Association of Public Health).

There are a growing number of education and training initiatives already underway. The International Development Research Council (IDRC) in Ottawa, for example, has contracted evaluators to prepare a developmental evaluation curriculum for local project managers in countries receiving bilateral development assistance from that organization. Similarly, the J.W. McConnell Family Foundation in Montreal has recently published ***Developmental Evaluation 201***, a manuscript that describes the experience of a group of evaluators employing developmental evaluation in a national project to promote youth engagement.

The approach that these bodies take to education and training should itself be developmental and continually adapt to new experience, learning's and perspectives in the field.

(c) gather, mine, distil and disseminate practice

Many of the participants in this study expressed a strong interest in learning more about the experiences and approaches employed by other people employing developmental evaluation. The areas that received the most frequent attention include: diagnostics for assessing the suitability of developmental situation, inquiry frameworks and methods that might be employed in developmental evaluation, and aids for budgeting, planning, monitoring and adjusting developmental evaluation in dynamic contexts.

The contingency-based approach to methodology in developmental evaluation means that the most effective way of documenting, learning from and refining its theory and practice is by documenting, sharing and analyzing actual case studies.

(d) facilitate connections between practitioners

Many study subjects reported an interest in connecting with other evaluation practitioners. This will create opportunities for evaluators to learn more about each other's approach to developmental evaluation as well as create opportunities for them to create networks amongst themselves. As noted earlier in this thesis, there already exist on-line communities of practice on the topic and spontaneous. Based on my own first-hand experience in one learning community devoted to the topic, I believe these mechanisms are excellent opportunities for learning and support.

6.4 Options for Further Research

This thesis was intended to be a preliminary investigation into the experience and reflections of early adopters of developmental evaluation as conceived by Patton. It is not surprising that in the process, it also uncovered other potentially productive lines of academic research.

6.4.1 Expand the Range of Inquiry frameworks

It would be useful for practitioners, researchers and theorists to gain a better understanding of the various types of inquiry frameworks employed in developmental evaluations. The aim of these investigations would not be to surface and refine possible (rather than best) frameworks and methods in developmental evaluation; one of the central ideas of

developmental evaluation is that the each evaluative situation is different and therefore requires a unique approach to evaluation.

More systematic data on different frameworks would help theorists and practitioners develop and test taxonomies that describe the conceptual and practical strengths and weaknesses of different approaches, and the conditions within which they thrive or wither. It is possible that such research might uncover patterns of inquiry and methods in different developmental situations (e.g. pre-formative, replication) and/or domain areas (e.g. education, health, workforce development), information that would help evaluators in different fields.

6.4.2 Explore the Fidelity of Developmental Evaluation Practice

Understanding the extent to which evaluators adhere to the ideal principles and operational practices of developmental evaluation would help evaluators understand where and how they might improve their practice of developmental evaluation. It would also provide signals to theorists such as Patton and others regarding how they might refine or adjust the theory, or simply acknowledge its inherent limitations. For example, in their review of forty-seven case studies of empowerment evaluation, Miller and Campbell (2006) concluded that few evaluators had consistently manifested all the ten principles of the approach. They concluded that the evaluation theorists needed to re-examine these principles and offer guidance to practitioners on how they might be better employed, particularly in projects of different sizes, scopes and aims.

It was too early to carry out this type of research in 2010. Patton's conception of developmental evaluation was still emerging and only a small number of evaluators had sufficient understanding of the existing literature and ideas on the approach to experiment with them in practice. It is worth stating again, therefore, that the aim of this research was to inform the development of developmental evaluation, rather than judge the fidelity of its practice.

The release of Patton's (2010) book on developmental evaluation will likely mean that many more evaluators will be able to experiment and use a more robust conception of developmental evaluation. Comparing the extent to which their actual application of

developmental evaluation aligns with the ideal and to explore the implications for modifying theory of practice are important future research objectives.

6.4.3 Examine the Challenge of Utilization

This study identified the critical challenge of use and mis-use of evaluation findings and process in the literature review on evaluation but did not explore it in interviews with participants. This is worth further investigation.

Evaluators have expressed a concern about the extent to which evaluation process and findings influence the thinking and decisions of the audiences of evaluation ever since the field gained legitimacy and began to grow in the 1960s (Weiss 1972). Since then, evaluating the use of evaluation has helped evaluators to determine if and how they might adjust their practice or theory to improve the probabilities that their efforts will be used to inform the thinking and decisions of the evaluation users.

While there already exists a great deal of primary, secondary and meta research on the factors related to evaluation use, targeted research on the use on the challenge of use developmental situations and developmental evaluation may reveal additional information on this matter.

6.5   Closing Remarks

The purpose of developmental evaluation is to assist people and organizations working in developmental settings to more systematically and rigorously use evaluative thinking and data in their work. The introduction of the theory or conception of developmental evaluation by Patton has added legitimacy to a diverse and eclectic range of practice that appears to have existed for some time and provides a tangible, well-researched, approach to inform practice in the field and the thinking and debate of evaluation theory.

The demand for developmental evaluation is likely to be strong as evaluators and evaluation users working in developmental situations seek out an evaluative orientation and practice suited to the unique nature of complex, fast moving and emergent situations. In a recent article, well- known Canadian evaluators argue that in order social innovation and rigorous

evaluation can find common ground in public institutions, government agencies must embrace evaluation as a process to encourage learning, reform and transformation, not simply accountability (Sridharan, Mayne & Nakamaia 2011).

In keeping with any new conception or methodology introduced over the past sixty years, the theory and practice of developmental evaluation will evolve. It will be continually (re)shaped by the repeated and diverse applications in real life settings as well as by continuing debates in the evaluation field (e.g. how to improve utilization of evaluation findings, to whom is an evaluator ultimately accountable, etc.).

Regardless of the extent to which they agree or do not agree with the conception of developmental evaluation, evaluators – and evaluation users who would like to productively employ evaluation in developmental situations – benefit from having one more distinct approach and approach they can draw upon to ensure that the field of evaluation is relevant, effective and responsive to an ever-changing world.

References

Adler, P.A & Adler P. 2001. The Reluctant Respondent. J.A. Gubrium & J.A.Holstein (Eds.), *Handbook of Interview Research: Context and Method*. Thousand Oaks, California: Sage Publications: 515-536.

Alkin, M. C. 1975. Evaluation: Who Needs It: Who Cares. *Studies in Educational Evaluation*. Volume 1. Number 3: 201-212.

Ariely, D. 2009. *Predictably Irrational: The Hidden Forces that Shape Our Decisions*. New York, New York: Harper Collins.

Arnstein, Sherry. 1969. A Ladder of Citizen Participation. Journal of the American Institute of Planners. Volume 35. Number 4: 215-224.

Auspos, P. and Kubisch, A. 2004. *Building Knowledge About Community Change: Moving Beyond Evaluations*. Washington, DC: The Aspen Institute Roundtable on Community Change.

Bamberger, M., Rugh, J. & Mabry, L. 2006. Real World Evaluation. Thousand Oaks, California: Sage Publications.

Berg, B. L. 2007. *Qualitative Research Methods for the Social Sciences*. Sixth Edition. Boston, Massachusetts: Pearson.

Bogdan, R. C., & Bilken, S. K. 2007. Qualitative Research in Education: An Introduction to Theory and Methods. Fifth Edition. Boston, MA: Allyn and Bacon.

Brofman, O. & Brofman, R. 2009. *Sway*. New York, NW: Broadway Business Books.

Campbell, S., & Fainstein, S.S., (Eds.). 2003. *Readings in Planning Theory*. Malden, Massachussetts: Blackwell Publishing.

Campbell, D.S., & Stanley, J.C. 1973. *Experimental and Quasi-Experimental Designs for Research*. Chicago, Illinois: Rand McNally

Carson, Mina. 1990. *Settlement Folk: Social Thought and the American Settlement Movement (1855 to 1930)*. Chicago, Illinois: Chicago University Press.

Chambers, Robert. 2010. *Paradigm, Poverty and Adaptive Pluralism*. Volume 2010, Number 344. Brighton, United Kingdom. Institute of Development Studies.

Chelimsky, E. 1997. The Coming Transformation in Evaluation. Editors. Chelimsky & W. Shadish. *Evaluation for the 21st Century*. Thousand Oaks, CA: Sage Publications.

Chen, H.T. 1996. A Comprehensive Typology for Program Evaluation. *Evaluation Practice*. Volume 17, Number 2: 121-130.

Chirban, J.T. .1996.. *Interviewing in Depth: The Interactive Relational Approach.* Thousand Oaks, CA. Sage Publications.

Colton, D. 1997. The Design of Evaluations for Continuous Quality Improvement. Evaluation and the Health Professions. Volume 20, Number 3: 265-85.

Crabtree, B.F., & Miller, W. L. (Eds.). 1992. Doing Qualitative Research: Multiple Strategies: Newbury Park, CA: Sage.

Cronbach, L.J. & Associates. 1980. *Toward Reform of Program Evaluation*. San Francisco, CA: Jossey-Boss.

Cunningham, K. & Ricks, M. 2004. Why Measure? Nonprofits Use Metrics to Show That They Are Efficient But What if Donors Don't Care? *Stanford Social Innovation Review*. Summer: 44-51.

Daly, K.J. 2007. *Qualitative Methods for Human and Family Development*. Thousand Oaks, CA: Sage Publications.

Davidoff, P. 1964. "Advocacy and Pluralism in Planning." *Journal of the American Institute of Planners*. Volume 31. Number 4: 544-55.

Deitchman, S. 1976. *The Best Laid Schemes: A Tale of Social Research and Bureaucracy.* Cambridge, Massachussetts: MIT Press.

Dewey, J., Montrose, B.E., Schroeter, D.C., Sullins, C.D., Mattox, J.R. 2008. Evaluator Competencies: What's Taught Versus What's Sought. *American Journal of Evaluation.* Volume 29, Number 3. September: 268-287.

Dewey, John 1922. *Democracy and Education: An Introduction to the Philosophy of Education*. New York: The MacMillan Company.

Diamond, Jared. 1997. *Guns, Germs and Steel: The Fate of Human Societies*. New York, New York: WW. Norton and Company.

Ely, M., Anzul, M., Friendman, T., & Diane, G. 1991. *Doing Qualitative Research: Circles within Circles*. London: Routledge Farmer.

Emerson, Jed. 2000. The Nature of Returns: A Social Capital Markets Inquiry into Elements of Investment and the Blended Value Proposition. *Social Enterprise Series No. 17*. Boston, Massachusetts: Harvard Business School.

Etzioni, A. 1967. Mixed Scanning: A Third Approach to Decision-Making. *Public Administration Review*. Volume 27: 385-392.

Fetterman D. M., Kaftarian, A.J., Wandersman, A. (Eds.). 1996. Empowerment Evaluation: Knowledge and Tools for Self-Assessment and Accountability. Thousand Oaks, California: Sage Publication.

Fetterman, D. 2000. Foundations of Empowerment Evaluation. Thousand Oaks, California: Sage Publication.

Fournier, D.M. 2005. Evaluation. *Encyclopedia of Evaluation*, edited by S. Mathison. Thousand Oaks, California: Sage Publications

Friedmann, J. 1966. "Planning as A Vocation". *Plan* (Canada). Volume 6, Number 2: 99-134.

_____1987. *Planning in the Public Domain: From Knowledge to Action*. Princeton, New Jersey: Princeton University Press.

_____ 1993. Toward a Non-Euclidian Mode of Planning. *Journal of the American Planning Association*. Volume 59. Number 4. Autumn: 482-85.

Freeman, H. E. 1977. Boundaries of the Evaluation Research Field. *Evaluation Studies Review*. Annual 2: 24-30.

Gamble, Jamie A. A. 2008. *A Developmental Evaluation Primer*. Montreal, Quebec: The J.W. McConnell Family Foundation.

Gauthier, B., Borys, S. Kishchuk, N, Roy, S. 2006. Evaluation Practice in Canada: Results of a National Survey. *Canadian Journal of Program Evaluation*. Volume 21. Number 3. Special Issue: 1-42.

Gladwell, Malcolme. 2005. *Blink: The Power of Thinking Without Thinking*. New York. Little, Brown & Company.

Gregory, Robert. 1995. Accountability, Responsibility and Corruption: Managing the Public Production Process. J. Boston (Ed.). *The State under Contract*. Wellington, New Zealand: Bridget Williams Book.

Greene, J. 1997. Evaluation as Advocacy. *Evaluation Practice*. Volume 18. Number 1: 25-35.

Greene, J. & Caracellli, V. (Ed.s). 1997. "Advances in Mixed Methods Evaluation: The Challenges and Benefit of Integrating Diverse Paradigms". *New Directions in Evaluation*. Number 74.

Guba, E. G. & Lincoln, Y.S. 1989. *Fourth Generation Evaluation*. Sage: San Francisco, California.

Gubrium, J.A, & Holstein, J.A. 2001. *Handbook of Interview Research: Context and Method*. Thousand Oaks, CA. Sage Publications.

Haag, K. and Rosengren, F. 2001. The Strong Focus on Output Information: A Threat to Evaluation in the Swedish State Sector? *Canadian Journal of Program Evaluation*. Volume 16. Number 1. Spring: 25-41.

Hall, Peter. 2002. Cities of Tomorrow: An Intellectual History of the Twentieth Century. London: Blackwell Publishing.

Halpern, Robert. 1994. *Rebuilding the Inner City: A history of Neighborhood Initiatives to Address Poverty in the United States*. New York, New York: Columbian University Press.

Haselkorn, Florence. 1978. Accountability in Clinical Practice. *Social Casework*. June. Volume 59: 330-336.

Hoisington, S.J. 1977. Accountability in Social Welfare. *Encyclopedia of Social Work.* National Association of Social Workers: 207.

House, E.R. 1990. Methodology and Justice. *Evaluation and Social Justice: Issues in Public Education. New Directors for Program Evaluation* 45: 23-36.

_____. 2004. Intellectual History in Evaluation. M.C. Alkin (Ed.). *Evaluation Roots: Tracing Theorists' Views and Influences*. San Francisco, California: Sage Publications.

House, E.R., & Howe K. R. 1999. Values in Evaluation and Social Research. Thousand Oak, California: Sage Publications.

Hum, D. & Simpson, W. 2002. Public Training Programs in Canada: A Meta Evaluation. *Canadian Journal of Program Evaluation*. Volume 17. Number 1: 119-138.

Jacobs, Jane. 1961. *The Death and Life of Great American Cities*. New York, New York: Random House.

Johnson, J.C. & Wells, S.C. 2001. Elicitation Techniques for Interviewing. J.A. Gubrium & J.A.Holstein (Eds.), *Handbook of Interview Research: Context and Method*. Thousand Oaks, California: Sage Publications: 491-514.

Kazi, M. 2000. Contemporary Perspectives in the Evaluation of Practice. *British Journal of Social Work*. Volume 30. Number 6: 755-768.

King, J.A. 1998. Making Sense of Participatory Evaluation in Practice. *New Directions in Evaluation*. Volume 80: 57-67.

Kirkendell, Richard. 1966. Social Science and Farm Politics in the Age of Roosevelt. Columbia, Missouri: University of Missouri Press.

Kluger, Jeffrey. 2008. *Simplexity: Why Simple Things Become Complex (and How Complex Things Can be Made Simple)*. New York, New York: Hyperion.

Kusek, J.Z. and Rist, R.C. 2002. *Building a Results-Based Monitoring and Evaluation System*. Washington, D.C: The World Bank.

Kramer, M. & and Bickel, W. 2004. Foundation and Evaluation As Uneasy Partners in Learning. M. Braverman, N. Constantine and J.K. Slater (Eds.) *Foundations and Evaluation: Contexts and Practices for Effective Philanthropy*. San Francisco: California: Jossey-Boss: 51-75

Krippendorf, K. 2008. *The Way of Innovation: Master the Five Elements of Change to Reinvent Your Products, Services, and Organization*. Avon, Massuchessetts: Platinum.

Kubisch, A., Auspos, P., Brown, P., Chaskin, R., Fulbright-Anderson, K. & Hamilton, R. 2002**.** *Voices from the Field II: Reflections on Comprehensive Community Change***.** Washington, DC: The Aspen Institute, Roundtable on Comprehensive Community Initiatives for Children and Families: Washington, District of Columbia.

Kvale, S. 1996. *Interviews: An Introduction to Qualitative Research Interviewing*. Thousand Oaks, California: Sage Publications.

Landry, R., Lamari, M. & Amara, N. 2003. The Extent and Determinants of the Utilization of University Research in Government Agencies. *Public Administration Review*. Volume 62. Number 2: 192-206.

Lincoln, Y.S. & Guba, E.G. 1985. *Naturalistic Inquiry*. Newbury Park, California: Sage Publications.

Lindblom, Charles E. 1959. "The Science of Muddling Through". *Public Administration Review*. Volume 39. Number 6: 79-88.

Mabry, L. 2010. Critical Social Theory Evaluation: Slaying the Dragon. *New Directions for Evaluation*. Special Edition. Critical Theory and Evaluation Practice. Volume 10. Issue 127. Autumn: 83-98.

Margoluis, R & Safalsky, N. 1998. *Measures of Success: Designing, Managing, and Monitoring Conservation and Development Projects*. Washington D.C.: Island Press.

Mark, M., Henry, G., & Julnes, G. 2000. *Evaluation: An Integrated Framework for Understanding, Guiding, and Improving Public and Non-Profit Policies and Programs*. San Francisco: Jossey-Boss.

Marshall, Catharine & Rossman, Gretchen B. 2011. *Designing Qualitative Research*. Los Angeles, California: Sage Publications.

Martin, Roger. 2007. *The Opposable Mind: How Successful Leaders Win Through Integrative Thinking*. Boston, Massachusetts: Harvard Business School Press.

_____. 2009. *The Design of Business*. Boston, Massachusetts: Harvard Business School Press.

Mayne, J. & Rist, R.C. 2006. Studies are Not Enough: The Necessary Transformation of Evaluation. *The Canadian Journal of Program Evaluation*. Volume 21. Number 3. Special Issue: 93-120.

Mintzberg, Henry. 1994. *The Fall and Rise of Strategic Planning*. New York, New York: Free Press.

Miles, M. B., & Huberman, A. M. 1984. *Qualitative Data Analysis*. Thousand Oaks, California: Sage Publications.

Millar, Ron Lin and Rebecca Campbell. Taking Stock of Empowerment Evaluation: An Empirical Review. *American Journal of Evaluation*. Volume 27. No. 3. September 2006: 296-319.

O'Conner, A. 1995. Evaluating Comprehensive Community Initiatives: A View From History. J.P. Connell, A.C., Kubisch, L.B. Schorr, C.H. Weiss. (Eds.) *New Approaches to Evaluating Community Initiatives: Concepts, Methods and Contexts*. Washington, D.C.: The Aspen Institute.

Owen, J.M. & Rogers, P. 1999. *Program Evaluation: Forms and Approaches*. Second Edition. London: Sage Publications.

Owen, J.M. 2004. Evaluation Forms: Toward an Inclusive Framework for Evaluation Practice. in *Evaluation Roots: Tracing Theorists' Views and Influences*. San Francisco, California: Sage Publications.

Pascale, R. T., Mark M., & Linda, G. 2000. *Surfacing the Edge of Chaos: The Laws of Nature and the New Laws of Business*. New York, New York: Three Rivers Press.

Patton, M.Q. 1994. Developmental Evaluation. *Evaluation Practice*. Volume 15. Number 4: 311-322.

_____. 1999. Organizational Development and Evaluation. *The Canadian Journal of Program Evaluation*. Special Issue: 93-114.

_____. 2002. *Qualitative Research and Evaluation Methods*. Sage Publications. Thousand Oaks, CA: Sage Publications.

_____. 2006. Evaluation for the Way We Work. *The Nonprofit Quarterly*. Vol. 13 (1): 28-33.

_____. 2008. *Utilization Focused Evaluation*. Fourth Edition. Thousand Oaks, CA: Sage
    Publications:

Pawson, R. & Tilley, N. 1997. *Realist Evaluation*. London, UK: Sage.

Perrin, B. 1998. Effective Use and Misuse of Performance Measurement. *American Journal of
    Evaluation*. Volume 19, Number 3: 367-79.

Peterson, J. 2003. *The Birth of City Planning in the United States, 1840-1917*. Baltimore,
    Maryland: John Hopkins University Press.

Petroski, H. 2006. *Success Through Failure: The Paradox of Design*. Princeton, New Jersey:
    Princeton University Press.

Phillips, S. & Karine, L. 2004. The Snakes and Ladders of Accountability: Contradictions
    Between Contracting and Collaboration for Canada's Voluntary Sector. *Canadian Public
    Administration.* Volume 47. Issue 4:451-474.

Preskill, H. & Torres, R.T. (Eds.). 2001. The Learning Dimension of Evaluation Use. *New
    Directions for Evaluation*. Volume 88. San Francisco, California: Jossey-Boss.

Pressman, J. L. and Wildavksy, A. 1983. *Implementation: How Great Expectation in
    Washington Are Dashed in Oakland; or, Why Its Amazing that Federal Programs Works
    at All This Being a Saga of the Economic Development Administration as Gold by Two
    Sympathetic Observers Who Seek to Build Morals on A Foundation of Ruined Hopes.*
    Berkely, California: University of California Press.

Robert Woods Johnston Foundation. 2008. *Qualitative Research Guidelines Project*.
    http://www.qualres.org/HomePhil-3514.html

Rogers, P. 2005. Accountability. *Encyclopedia of Evaluation.* S. Mathison (Ed.). Thousand
    Oaks, California: Sage Publications.

Rossi, P. H. 2004. My Views of Evaluation and Their Origins. M.C. Alkin (Ed.). *Evaluation
    Roots: Tracing Theorists' Views and Influences*. San Francisco, California: Sage
    Publications.

Rossi, P. H., Freeman, H. E. & Lipsey, M. W. 1999. *Evaluation: A Systematic Approach. Sixth
    Edition*. Thousand Oaks, California: Sage Publications.

Rossi, P.H. & Berk, R.A. 1981. An Overview of *Evaluation* Strategies and Procedures. *Human
    Organizations*. Number 40:287.

Rossman, G.B. & Rallis, S.F. 2003. Learning in the Field: An Introduction to Qualitative Research. Second Edition. Thousand Oaks, California: Sage Publications.

Schorr, L.B. 1988. Within Our Reach: Breaking the Cycle of Disadvantage. New York, New York: Anchor Books

_____ 1997. *Common Purpose: Strengthening Families and Neighborhoods to Rebuild America*. New York, New York: Double Day.

_____ 2009. *Realizing President Obama's Promise to Scale Up What Works to Fight Urban Poverty*. Washington, D.C.: Center for the Study of Social Policy

Scriven, M. 1967. The Methodology of Evaluation. Tyler, R.W., Gagne, M., & Scriven M. (Eds.). AERA Monograph Series on Curriculum Evaluation. Volume 1. Chicago, Illinois: Rand McNally.

_____. 1972. Prose and Cons About Goal-Free Evaluation. Evaluation Comment. The Journal Of Educational Evaluation. Volume 3. Number 4: 1-7

_____. 1991. Evaluation Thesaurus. Fourth Edition. Newbury Park, California: Sage Publications.

_____. 2003. Michael Scriven on the Differences Between Evaluation and Social Science Research. *The Evaluation Exchange*. Harvard Family Research Project. Volume 9, Number 4. Winter: 1.

_____. 2004. Reflections. M.C. Alkin (Ed.). Evaluation Roots: *Tracing Theorists' Views and Influences*. San Francisco, California: Sage Publications.

Seasons, Mark. 2001. Epilogue. *The Canadian Journal of Program Evaluation*. Special Issue 2001: 11-119.

_____. 2002. Evaluation and Municipal Urban Planning: Practice and Prospects. *The Canadian Journal of Program Evaluation.* Volume 17. Number 1: 43-71.

Seigart, D. & Brisolara (Eds). 2002. *Feminist Evaluation: Explorations and Experience*. New Directions for Evaluation. Number 96. Toronto, Ontario: Jossey-Boss.

Seilbeck-Bowen, K.A., Brisolara, S., Tischler, C., and Whitmore, E. 2002. Exploring Feminist Evaluation: The Ground From Which We Rise. New Directions for Evaluation. Number 96. Wiley Periodicals..Winter: 3-8.

Shadish, W. R., Cook, T.D. & Leviton, L.C. 1991. Foundation of Program Evaluation: Theories of Practice. Newbury Park, California: Sage Publications.

Shadish, W.R., Cook, T.D., & Campbel, D. T. 2002. Experimental and Quasi-Experimental Designs for Generalized Causal Inferences. Boston, Massachusetts: Houghton Mifflin.

Smith, M. F. 1989. Evaluability Assessment: A Practical Approach. Boston, Massachusetts: Kluwer.

Snow, D.R. & Buss, T.F. 2001. Development and the Role of Microcredit. *Policy Studies Journal.* Volume 29. Number 2: 296-307.

Stake, R. 1974. Program Evaluation, Particularly Responsive Evaluation. Reproduced in W. B. Dockrell & D. Hamilton (Eds), *Rethinking Educational Research.* 1980. London: Hodder and Stoughton.

Stouffer, Samuel A., Edward A. Suchman, Leland C. DeVinney, Shirley A. Star, and Robin M. Williams, Jr. 1949. Studies in Social Psychology in World War II: The American Soldier. Vol. 1, Adjustment During Army Life. Princeton: Princeton University Press.

Stufflebeam, D. L. & Webster, W.J. 1981. An Analysis of Alternative Approaches to Evaluation. *Evaluation Studies Review Annual.* Number 6: 70-85.

Stufflebeam, D. J. 2004. The 21st Century CIPP Model: Origins, Development, and Use. M.C. Alkin (Ed.). *Evaluation Roots: Tracing Theorists' Views and Influences.* San Francisco: Sage Publication.

Stufflebeam, D.L. 1968. The Use and Abuse of Evaluation in Title III. Theory Into Practice. Volume 6: 126-133.

Thaler, R. H. & Sunstein, C.R. 2009. *Nudge: Improving Decisions About Health, Wealth and Happiness.* New York, New York: Penguin Books.

The Aspen Institute's Roundtable on Comprehensive Community Initiatives for Children and Families. 1995. New Approaches to Evaluating Community Initiatives: Theory, Measurement, and Analysis. Volume 2. Editors: Karen Fulbright-Anderson, Anne C. Kubisch, James P. Connell. Washington, DC: The Aspen Institute.

The Aspen Institute Roundtable on Community Change. 1997a. *Voices from the Field: Learning From the Early Work of Comprehensive Community Initiatives. Roundtable on Comprehensive Community Initiatives for Children and Families.* Washington, DC: The Aspen Institute.

The Aspen Institute. 1997b. The Roundtable on Comprehensive Community Initiatives for Children and Families. *New Approaches to Evaluating Community Initiatives: Concepts, Methods, and Contexts.* Volume 1. Editors: Jame P. Connell, Anne C. Kubisch, Lisbeth B. Schoor, Carol H. Weiss. Washington, DC: The Aspen Institute.

The Aspen Institute's Roundtable on Community Change. 2005. *Structural Racism and Youth Development: Issues, Challenges and Implications.* Washington, DC: The Aspen Institute.

Todd, P.E. & K.I. Wolpin. 2006. Ex Ante Evaluation of a Social Program. *Working Paper -6-022*. Philadelphia, Pennsylvania: Penn Institute for Economic Research.

Underwood, T. 2006. Microcredit Research in the European Union (EU) 15 and European Economic Area (EEA). *European Dialogue No. 36*. European Microfinance Network: 23-32.

United States Agency for International Development (USAID) 1996. Preparing an Evaluation Scope of Work. *Performance and Evaluation TIPS*. Number 3. Washington, D.C.

United States Government Accountability Office. 1990. *Prospective Evaluation Methods: The Prospective Evaluation Synthesis*. Washington, D. C.: Program Evaluation and Methodology Division.

Vanderbilt, T. 2009. Traffic: Why We Drive the Way We Do. Toronto, Ontario: Vintage Books Canada.

Van Oudenhoven, N. & W. Rekha. 1998. Replicating Social Programmes: Approaches, Strategies and Conceptual Issues. *Management of Social Transformations Discussion Paper Series No. 19*. Paris, France: United Nations Educational, Scientific, and Cultural Organization.

Vainio-Mattilo, A. 1999. *Navigating Gender: A Framework and Tool for Participatory Development*. Helsinki, Finland: Ministry for Foreign Affairs, Department of International Development Cooperation.

Washington, R. O. 1965. *Program Evaluation In the Human Services*. Milwaukee, Wisconsin: University of Wisconsin, Milwaukee Center for the Study of Advanced Services.

Watson, K. 1994. A Review of Four Evaluations of CED Programs: What Have We Learned in Two Decades? J. Hudson & B. Galaway (Eds.). *Community Economic Development: Social and Economic Perspectives*. Toronto: Thompson Educational.

Weaver, L. & Cousins, J. B. 2004. Unpacking the Participatory Process. *Journal of Multidisciplinary Evaluation*. Volume 1:19-40

Weiss, C. 1972. Utilization of Evaluation: Towards Comparative Study. C.H. Weiss (Ed.), *Evaluating Action Programs*. Boston, MA. Allyn and Bacon: 318-326.

Weiss, C., & Bucuvalas, M. 1980. *Social Science Research and Decision-Making*. New York, NY: Russell Sage.

Weiss, C. 1991. Evaluation Research in the Political Context: Sixteen Years and Four Administration Later. M.W. McLaughlin & D.C Phillips (Eds.).*Evaluation and Education: At Quarter Century*. Chicago, Illinois: University of Chicago Press: 211-231.

_____. 1988, Evaluation (2nd Edition). Upper Saddle River, New Jersey: Prentice Hall.

Westley, F., Zimmerman, B. & Patton, M.Q. 2006. *Getting to Maybe: How the World is Changed*. Random House Canada: Toronto, Ontario.

Wholey, J. 1981. Using Evaluation to Improve Program Performance. *Evaluation Studies Review Annual*. Volume 6: 59-60.

_____. 1983. *Evaluation and Effective Public Management*. Boston, MA: Little Brown.

Wholey. J.S., Harry P. Hatry, & Newcomer, K.E. (Eds.). 1994. *Handbook of Practical Program Evaluation*. San Francisco, CA: Jossey-Boss.

Williams, W. and Evans, J.W. 1969. "The Politics of Evaluation: The Case of Head Start." **Annals of the American Academy of Political and Social Science.** 385. September: 118-132.

Williams B., & Imam, I.. (Editors). 2007. *Systems Concepts in Evaluation: An Anthology*. Edge Press & The American Evaluation Association.

Wolf, R.L. 1975. Trial by Jury: A New Evaluation Method. Phi Delta Kappan. Volume 53, Number 3: 185-87.

Appendix 1 - Information Consent Letter for Interview

Date

Dear (insert participant's name):

This letter is an invitation to consider participating in a study I am conducting as part of my Master's degree in the School of Planning in the Faculty of Environment at the University of Waterloo under the supervision of Professor Mark Seasons. I would like to provide you with more information about this project and what your involvement would entail if you decide to take part.

Michael Quinn Patton introduced the concept of 'developmental evaluation' (DE) – an approach used in the creation and/or ongoing adaptation of an intervention -- in a professional journal in 1994. Since then, he has published several more articles and provided approximately 72 workshops, presentations and interviews on the topic to thousands of people. He developed the idea of DE more fully in his latest edition of Utilization-Focused Evaluation (2008) and has written a book devoted specifically to the topic which is scheduled for release in July 2010.

Despite the number of years that the evaluation community has heard about DE, there is little information about how other evaluators beyond Mr. Patton understand and apply the concept of DE.

The purpose of this study, Developmental Evaluation: The Experience and Reflections of Early Experimenters, is to explore the experience and reflections of people who have used developmental evaluation in their work and to identify implications for refining and applying the concept in the evaluation community.   I would like to include you in my study because of your interest in, understanding of, and experience using DE. I have attached the questions I would to explore with you in an interview.

Participation in this study is voluntary. It will involve an interview of approximately 60 - 90 minutes in length to take place at mutually agreed upon time over the phone. You may decline to answer any of the interview questions if you so wish. Further, you may decide to withdraw from this study at any time without any negative consequences by advising the researcher. With your permission, the interview will be audio recorded to facilitate collection of information, and later transcribed for analysis. Shortly after the interview has been completed, I will send you a copy of the transcript to give you an opportunity to confirm the accuracy of our conversation and to add or clarify any points that you wish. All information you provide is considered completely confidential. Your name will not appear in any thesis or report resulting from this study, however, with your permission anonymous quotations may be used. Data collected during this study will be retained for three months after ht completion of the study in a locked office in my

supervisor's lab. Only researchers associated with this project will have access. There are no known or anticipated risks to you as a participant in this study.

If you have any questions regarding this study, or would like additional information to assist you in reaching a decision about participation, please contact me at 780-451-8984 or by email at mark@tamarackcommunity.ca. You can also contact my supervisor, Professor Mark Seasons at 519-888-4567 ext. 35922 or email mseasons@uwaterloo.ca.

I would like to assure you that this study has been reviewed and received ethics clearance through the Office of Research Ethics at the University of Waterloo. However, the final decision about participation is yours. If you have any comments or concerns resulting from your participation in this study, please contact Dr. Susan Sykes of this office at 519-888-4567 Ext. 36005 or ssykes@uwaterloo.ca.

I hope that the results of my study will be of benefit to those organizations directly involved in the study, other voluntary recreation organizations not directly involved in the study, as well as to the broader research community.

I very much look forward to speaking with you and thank you in advance for your assistance in this project.

Yours Sincerely,

Mark Cabaj

CONSENT FORM

By signing this consent form, you are not waiving your legal rights or releasing the investigator(s) or involved institution(s) from their legal and professional responsibilities.

I have read the information presented in the information letter about a study being conducted by **Mark Cabaj** of the Department of Environmental Resource Studies at the University of Waterloo. I have had the opportunity to ask any questions related to this study, to receive satisfactory answers to my questions, and any additional details I wanted.

I am aware that I have the option of allowing my interview to be audio recorded to ensure an accurate recording of my responses.

I am also aware that excerpts from the interview may be included in the thesis and/or publications to come from this research, with the understanding that the quotations will be anonymous.

I was informed that I may withdraw my consent at any time without penalty by advising the researcher.

This project has been reviewed by, and received ethics clearance through, the Office of Research Ethics at the University of Waterloo. I was informed that if I have any comments or concerns resulting from my participation in this study, I may contact the Director, Office of Research Ethics at 519-888-4567 ext. 36005.

With full knowledge of all foregoing, I agree, of my own free will, to participate in this study.

☐YES ☐NO

I agree to have my interview audio recorded.

☐YES ☐NO

I agree to the use of anonymous quotations in any thesis or publication that comes of this research.

☐YES ☐NO

Participant Name: _____ (Please print)

Participant Signature: _____

Witness Name: _____ (Please print)

Witness Signature: _____

Date: _____

Appendix 2 – Letter of Appreciation

Dear (*Name*);

I am writing to thank you for a stimulating meeting last week. It enjoyed learning more about your experience with and reflections on Developmental Evaluation.

My project, Developmental Evaluation: The Experience and Reflections of Early Experimenters, is unfolding according to design, and in particular my research for the chapter on the findings from this study is nearing completion. As you know I am now interviewing a few more individuals such as yourself who can lend additional information and insights and beginning to analyze the findings of the interviews.

I will send you a copy of the chapters that summarize my interviews with you and other interviewees. Please return the form to me by **[one week later]** through email so that I can proceed with the analysis.

I would like to confirm that I am on the only person who will review and analyze the results of this study. Your identity and responses will be coded in any written summaries of the interviews and in the final research report.

I hope you will get in touch with me if further thoughts occur to you about the subject of our conversation, particularly if you decide in retrospect that you would like to designate some of it for non-attribution. Should you have any comments or concerns you could also contact Dr. Susan Sykes of our Office of Research Ethics at 519-888-4567 Ext. 36005 or ssykes@uwaterloo.ca. This project was reviewed by, and received ethics clearance through, the Office of Research Ethics at the University of Waterloo.

I shall as promised, be sending you a typescript copy of the chapter, for your criticism and comments. I expect it to be ready for your review by October or November.

 Sincerely,

Mark Cabaj