OFF THE SIDES OF THEIR DESKS:
DEVOLVING EVALUATION TO NONPROFIT
AND GRASSROOTS ORGANIZATIONS

by

GERALD BRUCE HINBEST

M.A. Queen’s University, 1995
B.A. University of Guelph, 1979

A DISSERTATION SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF EDUCATION

in

THE FACULTY OF GRADUATE STUDIES

(Educational Leadership & Policy)

THE UNIVERSITY OF BRITISH COLUMBIA (VANCOUVER)

September 2008

© Gerald Bruce Hinbest, 2008
ABSTRACT

This study examines the changing context and implications for evaluation practice of social program and service delivery devolved to small nonprofit and grassroots organizations. The setting is explored through a critical reflection-on-practice of over twenty years experience conducting evaluation. Using a multiple case study approach, the dissertation examines nine broad themes through two broad composite scenarios and twenty-five detailed vignettes that portray the challenges of working as a consultant with and for small nonprofit and grassroots organizations as they grapple with growing demands for accountability through evaluation.

The multiple case study analysis is complemented by an analysis of case studies in two broad areas of literature; one on the impacts of devolution in the nonprofit sector, and the other examining recent trends in evaluation conducted in challenging settings, including community-based and non-governmental organizations (NGOs). The five broad themes addressed through the case studies and literature on devolution are: 1) accountability, 2) capacity, 3) mandate drift, 4) competition, and 5) complexity. The four broad themes addressed through case studies and literature on evaluation are: 1) theory-based evaluation, 2) inclusiveness (participatory approaches), 3) the changing and multiple roles of evaluators, and 4) the use of dialogue, deliberative and democratic approaches in evaluation practice.

The study contends that the ‘rough ground’ of nonprofit settings provides a useful lens for understanding broader challenges and trends in evaluation practice; that evaluators provide more than just technical skills and knowledge, but undertake important roles in linking communities, mediating among stakeholders, fostering dialogue and deliberation about programming, and mitigating some of the more egregious impacts of devolution experienced by nonprofit and grassroots organizations. By acknowledging and supporting the development of such roles and responsibilities, the profession and evaluators working in these settings can provide meaningful contributions to public discourse about the nature of accountability, the broad context of social programming, the complex capacity challenges being faced by nonprofit organizations, and the role of evaluation in exacerbating or potentially mitigating such effects.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABSTRACT</td>
<td>ii</td>
</tr>
<tr>
<td>TABLE OF CONTENTS</td>
<td>iii</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>v</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>vi</td>
</tr>
<tr>
<td>LIST OF VIGNETTES</td>
<td>vii</td>
</tr>
<tr>
<td>ACKNOWLEDGEMENTS</td>
<td>viii</td>
</tr>
<tr>
<td>CHAPTER ONE: INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>The Changing Context</td>
<td>2</td>
</tr>
<tr>
<td>A Reflection on Practice</td>
<td>6</td>
</tr>
<tr>
<td>Reflective Questions</td>
<td>7</td>
</tr>
<tr>
<td>Outline of the Dissertation</td>
<td>7</td>
</tr>
<tr>
<td>CHAPTER TWO: RESEARCH METHODS: REFLECTION ON PRACTICE</td>
<td>10</td>
</tr>
<tr>
<td>Study Rationale</td>
<td>12</td>
</tr>
<tr>
<td>Developing the Themes</td>
<td>14</td>
</tr>
<tr>
<td>Power and Reflexivity</td>
<td>15</td>
</tr>
<tr>
<td>A Multiple Case Study Approach</td>
<td>17</td>
</tr>
<tr>
<td>Multiple Case Studies</td>
<td>17</td>
</tr>
<tr>
<td>Data for the Case Studies</td>
<td>19</td>
</tr>
<tr>
<td>Triangulation</td>
<td>20</td>
</tr>
<tr>
<td>Composite Scenarios</td>
<td>21</td>
</tr>
<tr>
<td>Composite Scenario Background</td>
<td>25</td>
</tr>
<tr>
<td>Vignette 2.1 – The Grassroots Composite Scenario</td>
<td>25</td>
</tr>
<tr>
<td>Vignette 2.2 – The Small Nonprofit Composite Scenario</td>
<td>27</td>
</tr>
<tr>
<td>CHAPTER THREE: TRENDS IN THE NONPROFIT, VOLUNTARY AND GRASSROOTS SECTOR</td>
<td>30</td>
</tr>
<tr>
<td>Defining the Nonprofit Sector</td>
<td>31</td>
</tr>
<tr>
<td>Canadian Nonprofit Organizations</td>
<td>33</td>
</tr>
<tr>
<td>Civil Society</td>
<td>37</td>
</tr>
<tr>
<td>The Devolution of Programs and Services</td>
<td>40</td>
</tr>
<tr>
<td>The Rationale for Devolution</td>
<td>41</td>
</tr>
<tr>
<td>The Impacts of Devolution on Nonprofit Organizations</td>
<td>47</td>
</tr>
<tr>
<td>CHAPTER FOUR: THE ROLE OF EVALUATION IN SOCIETY</td>
<td>49</td>
</tr>
<tr>
<td>Devolving Evaluation</td>
<td>49</td>
</tr>
<tr>
<td>Defining Accountability</td>
<td>57</td>
</tr>
<tr>
<td>Reflecting on Evaluation and its Rationale</td>
<td>65</td>
</tr>
</tbody>
</table>
CHAPTER FIVE: CASE STUDY ANALYSIS: DEVOLUTION THEMES .............................................. 71

Theme One: Accountability ...................................................................................................... 72
  Accountability in the Context of Devolution ....................................................................... 72
  Accountability and Performance Measurement ..................................................................... 81

Theme Two: Capacity .............................................................................................................. 89

Theme Three: Mandate Drift .................................................................................................. 107

Theme Four: Competition ....................................................................................................... 127

Theme Five: Complexity ......................................................................................................... 145

Summary ................................................................................................................................. 167

CHAPTER SIX: CASE STUDY ANALYSIS: EVALUATION THEMES ....................................... 174

Theme Six: Theory-Based Evaluation ..................................................................................... 175

Theme Seven: Inclusiveness .................................................................................................... 203

Theme Eight: Roles ................................................................................................................ 231

Theme Nine: Dialogue, Deliberation & Democracy ................................................................ 261

Summary ................................................................................................................................. 282

CHAPTER SEVEN: OFF THE SIDES OF THEIR DESKS ......................................................... 285

Revisiting the Reflective Questions ....................................................................................... 287

Implications for Practice ....................................................................................................... 291

Leadership in Evaluation Practice ....................................................................................... 304

Distinctive Contribution of the Study .................................................................................. 307

Study Limitations ................................................................................................................ 308

Future Research .................................................................................................................. 308

Reflections on Researching my Practice ............................................................................... 311

REFERENCE LIST .................................................................................................................. 313

APPENDIX A: META-ANALYSIS TABLES ............................................................................. 339

APPENDIX B: CANADIAN NONPROFIT & VOLUNTARY ORGANIZATIONS ..................... 341
LIST OF TABLES

TABLE A1: DEVOLUTION CASE STUDIES AND SURVEY ARTICLES .............................................................. 339
TABLE A2: EVALUATION CASE STUDIES AND SURVEY ARTICLES .......................................................... 340
TABLE B1: NONPROFIT & VOLUNTARY ORGANIZATIONS, BY PRIMARY ACTIVITY AREA, 2003 ........ 341
LIST OF FIGURES

FIGURE 6.2.1: PROCESS DIMENSIONS IN COLLABORATIVE INQUIRY ........................................... 205
LIST OF VIGNETTES

VIGNETTE 2.1 – THE GRASSROOTS COMPOSITE SCENARIO ........................................................... 25
VIGNETTE 2.2 – THE SMALL NONPROFIT COMPOSITE SCENARIO ........................................... 27
VIGNETTE 5.1.1 (NONPROFIT) – OSSIFYING LOGICS ................................................................. 77
VIGNETTE 5.1.2 (GRASSROOTS) – ELEVENTH-HOUR REFRAMING ............................................. 83
VIGNETTE 5.2.1 (GRASSROOTS) – OFF THE SIDE OF THE DESK ................................................... 93
VIGNETTE 5.2.2 (NONPROFIT) – TRADE OFF ............................................................................ 98
VIGNETTE 5.3.1 (NONPROFIT) – THE NEW INFORMATION SYSTEM ........................................... 108
VIGNETTE 5.3.2 (GRASSROOTS) – STEALTHY CHANGES ........................................................... 116
VIGNETTE 5.4.1 (GRASSROOTS) – DECIDING TO GROW ........................................................... 129
VIGNETTE 5.4.2 (NONPROFIT) – ANTICIPATING OPPORTUNITIES ............................................ 137
VIGNETTE 5.5.1 (NONPROFIT) – TWO BIRDS WITH ONE ......................................................... 149
VIGNETTE 5.5.2 (GRASSROOTS) – SWITCHING HATS ............................................................... 153
VIGNETTE 5.5.3 (NONPROFIT) – REFRAMING GOALS & ACTIVITIES ........................................ 157
VIGNETTE 5.5.4 (GRASSROOTS) – REFRAMING SUCCESS & FAILURE .................................... 160
VIGNETTE 6.1.1 (NONPROFIT) – BUILDING A LOGIC ................................................................. 184
VIGNETTE 6.1.2 (NONPROFIT) – REFRAMING PROGRAM THEORY ........................................... 192
VIGNETTE 6.1.3 (GRASSROOTS) – SHARING A COMMUNITY OF PRACTICE .............................. 196
VIGNETTE 6.2.1 (NONPROFIT) – PARTICIPATION REVOKED .................................................... 206
VIGNETTE 6.2.2 (GRASSROOTS) – EVALUATOR OR PARTICIPANT ... OR BOTH? ...................... 212
VIGNETTE 6.2.3 (NONPROFIT) – REVOLVING MEMBERSHIP .................................................... 217
VIGNETTE 6.2.4 (NONPROFIT) – BUILDING LONG-TERM CAPACITY .................................... 222
VIGNETTE 6.3.1 (GRASSROOTS) – EVALUATION VS. RESEARCH ............................................. 232
VIGNETTE 6.3.2 (NONPROFIT) – CHOOSING SIDES ... OR NOT? ............................................. 241
VIGNETTE 6.4.1 (NONPROFIT) – ENGAGING DISCUSSION, DEBATE & DIALOGUE ................ 264
VIGNETTE 6.4.2 (GRASSROOTS & NONPROFIT) – COMMUNITIES OF DIFFERENCE ............ 275
ACKNOWLEDGEMENTS

I have been fortunate to make this dissertation journey with a particularly well-suited and supportive committee. I am indebted to Carolyn Shields for challenging me to critically explore my evaluation practice as a context of both education and leadership. Sandra Mathison’s breadth of knowledge about evaluation and her professional efforts to reflect on its basic assumptions gave me inspiration and practical insight that can be found woven throughout the fabric of the dissertation. I am particularly grateful to Tom Sork, who as chair of my committee demonstrated outstanding scholarship, gentle and unfailing encouragement, and exceptional patience with my seemingly interminable process. To all three of you, thank you for believing in this work as both worthwhile and achievable.

The Ed.D. Program is pointedly about examining and renewing professional practice. Key to this process is the learning community represented by the cohort, who provide practical wisdom from practice contexts, and model in word and deed their own growth and critical reflection. I thank all members of my cohort for their substantial and ongoing support and camaraderie, and I remain particularly indebted to Rhonda Margolis, Joanna Ashworth, and Jennifer White for their inspiration, sage counsel, and companionship. I also thank David Coulter and Shauna Butterwick, who so ably represent the values of the program.

I hope that this dissertation faithfully represents the serious challenges faced by nonprofit and grassroots organizations that I came to appreciate through informative discussions with project coordinators, administrators, and participants. My thanks in particular go to Jennifer Kyffin and Jennifer White, who proved to be wellsprings of innovative ideas and exceptional sounding boards for constructive reflection.

I extend my deep appreciation for valuable comments, clarifications, resource suggestions, and encouragement generously offered via e-mail by Ernie House, Bob Stake, Jennifer Greene, Bob Williams, Michael Patton, Jennifer Clark, Angela Eikenberry, Dvora Yanow, Susan Phillips, John Shields, Ted Richmond and Bryan Evans. Colleagues near and far have contributed substantially to my understanding of issues and my process of working through how to represent them, including Lynette Harper, Doug Fraser, Chantelle Marlor, Caroline Burnley, Owen Peer, Victor Owen and John Vellacott. Linda Derksen provided all this and more, creatively walking the line between department chair, motivator and friend.

The practical matter of writing the dissertation was greatly enhanced by delightful opportunities to house sit for Lynne MacFadgen & Ross MacKay, Joe & Rony Gerritsen, Lynette Harper & Bruce Finlayson, and Leigh Blaney. When not borrowing a desk from these friends, I was writing at a table in the Teahouse on Rutherford, for which I must thank the incomparable hospitality and enthusiasm of Bernice Rambo and Jaz Parmajit.

Two mentors have enriched my professional life. From David McKinney I learned about balancing rigour in research with a passion for social justice. Mike Nelson taught me how to grow beyond doing evaluation, and to question what it could and should represent.

It is beyond my means or the space available to acknowledge the many friends who offered me sustenance and understanding as I neglected other priorities. Please be aware that for your questions, accommodations and encouragements I remain profoundly grateful.

Last, but in my heart and mind, first, I thank my family. To my daughter Maggie, thank you for your enthusiasm, love and humour, and for your tremendous help with the references. And to Jeanne, whose patience, love and wisdom sustained me and reminded me of my purpose: without your enduring support this dissertation would not have been possible.
CHAPTER ONE:
INTRODUCTION

Evaluators are drawing closer to community, making their services available there. Perhaps this is a symptom of the growing distance between western governments and their electorates, a reaction by communities and professions against centrally-imposed, coercive programmes – an attempt, perhaps, to recapture the change initiative for the grass-roots... there seems to be a proliferation of, often small-scale, studies evaluating community development projects.

Saville Kushner, 2002b, p. 20.

The phone rings. It's Laura, the director of a local grassroots organization. I've been working with her on an evaluation of one of her agency’s innovative pilot programs, and she has been reviewing the first draft of the final report I submitted before sharing it with her steering committee. Her somewhat panicked voice says, “But where are the numbers”?

Laura's question is not about the appropriateness of qualitative versus quantitative inquiry; it is concerned with the nature of accountability in delivering programs and services in small nonprofit organizations. It turns out that at a meeting earlier that day on an unrelated issue, the funder’s representative had offered congratulations on completing the project, and expressed anticipation for receiving the report and “seeing how the numbers turned out.”

Despite spending over a year working closely with Laura and a broadly-based steering committee in jointly designing, developing and implementing the project, and working out detailed data collection plans since the project’s initial planning stages, one casual comment to her by a representative of the funder as the initiative nears completion is enough have her questioning our work and progress.

While I am a little taken aback by Laura's phone call, I am not completely taken by surprise. It is a phone call I have had repeated on numerous occasions. Even with programs that have had only a dozen participants over a year-long period, and for which the primary
data collection approaches involved detailed interviews, focus groups, observation or a case study, at some point the perceived needs of accountability to a funder can quickly set aside all previous planning, discussions and understandings. As I work to reassure Laura that we can and should stick to our game plan, that the funder will be satisfied with the report, and that her board of directors will appreciate this form of reporting, I am reminded that the negotiation and dialogue never end – they start before the project begins and may not finish until long after the project is completed.

In this dissertation I examine the changing context and implications for evaluation in small nonprofit and grassroots organizations. The topic is program evaluation, but my focus is not the methods of research, but rather, the socio-political context in which program evaluation takes place. The dissertation represents a critical reflection-on-practice of some twenty-five years spent conducting program evaluation, as an internal evaluator within government, and as an external consultant to all levels of government and to nonprofit organizations. The dissertation grew out of my ongoing efforts to understand the sources of change I have experienced in my practice, and to understand the implications of my role as an evaluator for the organizations I have worked for and with.

**The Changing Context**

Changes to the program and service context have been profound. Over the past two decades, the New Public Management (NPM) movement (Kaboolian, 1998) has been associated with a wide range of reforms to the relationship between state, the private sector and the broader public sector across much of the Western world. Two aspects of these changes have proven to be particularly meaningful for the nonprofit sector – the general downsizing of government, and the concomitant shifting of responsibilities for program and
service delivery to the nonprofit sector. This shifting of program and service delivery has been widely characterized as ‘devolution,’ however it encompasses a diverse range of descriptive and analytical distinctions. The language used to discuss ‘devolution’ reflects the multi-disciplinary ramifications of the phenomenon. The terminology is both inconsistent and evolving. The general discourse has been theoretical; the political science and sociological as well as other academic disciplines and literatures have used labels such as neoliberal and neoconservative transitions, marketization, privatization, and decentralization. In the literature of human services and social work practitioners, the discussions have tended to be more practically based – speaking of devolution, downsizing and outsourcing. The administrative disciplines have used the full range of terms, depending on the publication and the intended audience. All approaches talk about the issues in terms of the ‘third, voluntary, nonprofit or grassroots’ sectors. More recently, the theoretical discussions have focused on ‘civil society’ as a frame for conceptualizing the sector. The terminology itself is an important facet of understanding and framing the issues faced by the sector, and will be discussed more fully in Chapter Three.

The devolution of programs and services to the nonprofit and voluntary sectors has generally involved organizations adopting many of the methods and values of the market. The New Public Management has primarily been characterized by an increasingly hierarchical approach to governance, such as contractual and competitive project-based funding and a preoccupation with accountability. However, recent reforms have attempted to introduce what purport to be more horizontal forms of governance processes – working with and through networks, using cooperative approaches of decision-making and policy development, and coordinating among diverse stakeholder groups. Recent discussions have
highlighted an apparent shift toward more horizontal governance models (Levasseur & Phillips, 2005; Phillips, 2004; Salamon, 2003). My interest lies in the tensions that exist and are emerging between accountability on one hand, and autonomy, collaboration and interdependence on the other – in short, the nexus of governance within nonprofit, voluntary and grassroots organizations. To date, the accountability emphasis in the hierarchical governance under New Public Management has been on contract compliance, control of abuse, and the appropriate use of public funds. Although this is only one aspect of the broader mandate of evaluation, the demand for performance improvement and effectiveness evaluation has been growing, and there is a real need to understand how processes of devolution affect those working on the front lines of evaluation and accountability in the nonprofit sector.

As program and service delivery has devolved to nonprofit and voluntary organizations, the locus and nature of much evaluation work has also changed. Devolution has broad implications for the evaluation field, reflecting shifts in the process, content, clients and stakeholders of evaluation. Similarly, the increased emphasis on public accountability associated with the devolution of programs and services is reflected in changes to how programs are designed, administered and funded, and to the kinds of programming available to clients. And yet the evaluation literature has been curiously silent on the broader impacts and implications of devolution for both evaluation theory and practice.

On the other hand, the past few years have seen many of devolution’s impacts on nonprofit organizations documented – mission drift, a reduced advocacy role for organizations and their clients, increased competition among agencies, and planning challenges because of the contractual nature of funding (Evans & Shields, 2005; Eikenberry
& Kluver, 2004; Richmond & Shields, 2004; Phillips & Levasseur, 2004). Such changes are fundamentally implicated in what evaluators do.

The quote by Saville Kushner (2002b) that introduced this chapter, which spoke of evaluators drawing closer to community, and the proliferation of small scale studies evaluating community development projects, continued as follows:

Here, the evaluator has to make no effort to discover desirable agendas. These are often values-driven projects, which grow directly out of need and are neither mediated through the logic of government nor their ethical purposes overwhelmed by distantly-conceived targets. They are often more or less pure well-springs of humanism. (p. 20)

The experience from my practice on which I reflect in this dissertation suggests to me that Kushner’s optimism here is unwarranted. While such projects at the grassroots may begin as values-driven projects, they clearly do quickly become drawn into the logic of government as soon as they begin the process of applying for or requesting financial support from the state – even if they never obtain that funding.¹ This begins with how the project is framed or re-framed in initial proposals, and continues through funding processes that demand complex accountability of even the most small and innovative projects, and through the ongoing reshaping of community initiatives to make them appear attractive for renewed funding or to new or additional funders. The system of contracting and financial supports under the New Public Management can quickly transform ‘well-springs of humanism,’ redefining and repackaging innovative ideas, and removing from them their contingent, situated, and locally relevant emphasis. In the process, organizations are drawn into the world of contracted program delivery, and refashion their identities within the community and beyond.

¹ Indeed, it begins long before that, based on how the state defines problems, allocates funding, and decides what is within the scope of government intervention.
A Reflection on Practice

The University of British Columbia’s Doctor of Education in Educational Leadership and Policy (Ed.D.) program uses reflection on practice to help students examine root assumptions about the contexts in which we work. Such pointed reflexivity challenges our practices and enriches them. I enrolled in the Ed.D. Program because I had questions and concerns about my ongoing practice as a consultant and evaluator, and I was searching for tools to help me understand the implications of that practice. The Ed.D. is specifically designed to produce knowledge that contributes to the understanding and improvement of practice, in addition to a broader theoretical understanding of practice issues.

The dominant theme in my practice over the past twenty-five years has been program evaluation. My evaluation practice has been in federal and provincial government bureaucracies (as an internal evaluator), with two small management-consulting firms, as an independent consultant, in post-secondary institutions, and as a volunteer with community-level nonprofit organizations. Over this period I participated in or was the lead evaluator for approximately fifty program evaluation studies, spanning large national programs to small innovative projects offered by grassroots organizations, which involved perhaps a dozen participants. Over time, my focus shifted from large-scale programs to smaller provincial level programs, and finally, over the past dozen years, to programs delivered in local communities by third party deliverers. For some time I had seen these changes as reflecting my own shifting interests and ongoing frustrations with the contexts in which I have worked. Through the reflection on practice afforded by the Ed.D. program, I have come to appreciate how these changes represent in at least equal measure transformations occurring in the broader context of evaluation work.
In this dissertation I reflect on these changes, describe and document the range of impacts on organizations and clients of the process of devolution to third party deliverers, and also identify challenges and implications of the devolution of evaluation activities to local communities. In doing so, I draw upon my experience as a program evaluator working with a variety of organizations in many contexts. While the story I am telling involves the experience and insights of many people – collaborators and clients from my practice over the past dozen years, in particular – the reflection is of my experience as an evaluator in these contexts. It is an on-the-ground examination of devolved evaluation practice.

Reflective Questions

My reflection on practice in this dissertation focuses on the following questions:

1. How and why has my evaluation practice changed over the past twenty-five years?
2. How does the devolution of programs and services affect the practice of evaluation?
3. What are the problems and challenges of conducting evaluation in a context of devolved programs and services?
4. How does evaluation affect devolved programs and services?
5. What are the implications of devolving evaluation to nonprofit organizations for clients, programs, organizations, communities, evaluators and the broader society?

The questions as posed reflect a dual focus – the nature of devolution and its impacts on nonprofit and grassroots sector organizations, and the nature of evaluation within that context. My overriding interest is in the implications for evaluation practice, but it is the devolution context that marks its academic and practical significance.

Outline of the Dissertation

Chapter Two further describes the rationale for the approach taken for this reflection on practice. It presents the idea of a ‘multiple case study’ analysis for examining the diverse
contexts and examples from my practice over the past dozen years. The chapter introduces and describes the broad experience base on which I reflect, the presentation format for the ‘vignettes’ through two composite scenarios, and the broad literatures used to help identify the analytic themes that I use to frame the analysis and presentation of arguments.

Chapter Three summarizes the literature examining trends in the Canadian nonprofit, voluntary and grassroots sector. The two main sections of the chapter examine the current profile of nonprofit organizations in Canada, and some of the history of devolution of programs and services to nonprofit organizations, respectively. Because many of the examples and case studies in the literature reflect the nonprofit context in Europe, Southeast Asia, and the United States, this chapter also discusses devolution and nonprofit organizations in an international context, and reflects on broad global trends and recent discourse concerning the relevance and importance of the concept of civil society to the sector.

Chapter Four describes some of the changes to the contexts in which evaluation work is done, and reviews key trends in evaluation, assessing the relevance of these trends for the context of devolution as a site of evaluation activity. The devolution process involves shifts in who is conducting evaluation, for which programs, and for whom, as the clients of evaluation are becoming more diverse and simultaneously less easily identifiable. Such trends to become more democratic, deliberative, participatory, inclusive, developmental, and responsive reflect an acknowledgement of the inherent complexity and diversity of evaluation contexts. The chapter situates my own experience and understanding of evaluation within the literature by focusing on my evaluation practice from the perspective of on-the-ground constructed systemic relationships among community players.
Chapter Five provides a structured approach to understanding the impacts of devolution on nonprofit organizations and other stakeholders, and the impacts and implications for program evaluation as well. The chapter focuses on five devolution themes – accountability, capacity, mandate drift, competition and complexity. These themes are constructed through examining the case studies drawn from my practice and from analysis of the literature on devolution.

Chapter Six represents the evaluation counterpart to Chapter Five. It provides a structure for reflecting on four broad themes from the evaluation literature, and how they are reflected in and relevant to devolved evaluation contexts. The chapter focuses on the themes of theory-based evaluation, inclusiveness, the multiple roles of the evaluator, and dialogue, deliberation and democracy as both analytical approach and practical strategy for on-the-ground practice.

In Chapter Seven I reflect on the implications of the nine themes described and discussed in Chapters Five and Six. The chapter summarizes key arguments, explores the implications for devolved program and service delivery and for evaluation in a devolved context, and reflects on the relevance of this context to the broader world of evaluation practice. Finally, the chapter suggests a framework for future research and reflective practice.
The single case is meaningful, to some extent, in terms of other cases. The researcher and the readers of the case report are acquainted with other cases. Any case would be incomprehensible if other, somewhat similar cases were not already known. So even when there is no attempt to be comparative, the single case is studied with attention to other cases.


The literatures dealing with evaluation practice and with service delivery and administration regularly use case studies to demonstrate practice issues. In examining my practice of evaluation, the issues I am most interested in articulating and examining are those that have not been discussed widely in the literature. While Khakee (2003) has spoken of the growing gap between theory and practice in evaluation, and Stame (2006) notes the divergence between approaches that attempt to simplify contexts (performance measurement) versus those that acknowledge and address complexity (realistic evaluation, developmental evaluation, responsive evaluation, fourth generation evaluation), there are other gaps in the literature that have not been examined analytically. Most discussions in both the administrative / human service and evaluation literatures focus on larger organizations. When nonprofit and voluntary organizations are mentioned, it is typically larger foundations and umbrella organizations that are the focus – little has been done on the grassroots. As I have examined the evaluation literature that has been produced over the past fifteen years, I have noted that few articles – even those that do address the reality of evaluation in nonprofit organizations – examine grassroots as a context of practice, and fewer still look at the implications of the devolution of programs and services for evaluation practice. Similarly, while the literature on social program and human service delivery does address nonprofit organizations, and increasingly examines issues related to devolution, its emphasis is still on
larger nonprofit and voluntary organizations, leading Smith (1997) to refer to the grassroots as “the dark matter ignored in prevailing ‘flat earth’ maps of the sector” as he notes how seldom studied such groups have been.

The complexity of the evaluation contexts that I want to understand and capture is challenged by traditional approaches to the case study. No one case could provide the opportunity to examine the breadth of inter-related issues that are encompassed by devolution as a context for practice. It is this complexity that I am interested in exploring, for it is in the combination and interaction of issues that I believe it is important to reflect on the implications for communities, organizations and evaluation practice.

To address this combination of breadth and complexity I use a multiple case study (Stake, 2006) approach to conducting my reflection on practice. While the review is retrospective, and does not involve new collection of data concerning the various sites of my practice over the past dozen years, it does offer a diverse combination of settings on which to reflect. Over the past dozen years I have worked primarily in consulting – most of that time as an independent consultant. During this time I undertook approximately twenty-five projects, of which sixteen were evaluations, and seven involved nonprofit organizations directly as clients. Three evaluation projects were undertaken on behalf of federal government departments, three for provincial government departments, and five for universities or research centres working through universities. All of the evaluation work done for the provincial and university sectors involved working with third-party delivers of programs and services, as did at least another half dozen evaluation studies undertaken previously as an internal evaluator working within provincial government departments. While I refer to the latter examples on occasion, the main focus of my review of practice
consists of the seven evaluations with nonprofit clients, and the eight evaluation studies undertaken for provincial departments or universities, all of which were done in the period between 1995 and 2006.

**Study Rationale**

The literature on devolution in Canada has matured over the past ten years. It has begun to move beyond the focus on describing and documenting the nature of devolution, and has provided extensive evidence of potentially devastating impacts of that devolution. A key element of this impact has been the New Public Management emphasis on accountability, and this emphasis has frequently been translated and expressed in the form of expectations for project or program evaluation. Further, the evaluation process itself is being devolved to nonprofit, voluntary and grassroots organizations, demanding new skills, knowledge and competence of those working in the sector. While larger non-profits sometimes have staff that are prepared and capable of undertaking such tasks, the majority of organizations do not (Carman, 2005 & 2007; Hall et al., 2003; Fine et al., 1998). In some cases, the organizations hire consultants to help with such work. At other times, particularly in small organizations, the task falls to program managers, coordinators or front-line deliverers. Increasingly, it is some combination of these.

In my practice as a consultant, and to a lesser extent before that as an internal evaluator in government, I have seen this process emerging over the past twenty years. And yet there is little evidence of this trend in the evaluation literature. In part, this reflects the professional focus on evaluation of larger, long-term and established programs. Many long-term programs are still evaluated in this manner, although the changes to the context of delivery, even for large programs, has implications for how evaluation is conducted and
coordinated – reflecting a much more complex operating environment – and yet these changes have had only a minor influence on larger theoretical and practical debates in the evaluation literature.

My starting point in the dissertation is to document some of the changes that have transformed the context in which community level evaluation of programs and services is happening. This builds on insights from the literature and case studies that have documented the impacts of devolution on programs and services, and helps build a frame for reflecting on case examples from my own practice over the past dozen years. I then use some of the most recent trends in evaluation practice and theory to frame additional case studies from the evaluation literature, again building on the examples of devolved program delivery from my own practice. Finally, I use these conceptual tools to reflect on the implications for nonprofit organizations, their staff and clients, communities and evaluators working in a context of devolved evaluation.

While the reflection on practice in the dissertation is presented in the form of a multiple case study, a significant component of my research and analysis consists of a comprehensive meta-analysis of case studies (Jensen & Rodgers, 2001) and other literature related to the devolution of programs and services, evaluation in nonprofit and local community contexts, and the intersection of the two. Although I did not undertake new interviews or data collection for the study, the case study format provides a useful frame for presenting examples, and comparing them to other cases and studies that have been published over the past ten years. The latter task involved examining diverse studies in both the public administration / human service fields, and the literature of evaluation practice. Occasionally, they met in articles that attempted to address both sets of issues, although this was rare.
Developing the Themes

The starting point for my reflection on practice was examining and questioning changes occurring in the context of my work as an evaluator. I treated the specific evaluation projects on which I worked as individual cases that demonstrated some constellation of the factors and themes that I saw emerging. I extracted five broad devolution themes and four evaluation themes from the case studies, using the meta-analysis as a guide to help frame the issues and define themes concerning devolution and evaluation within that context. These themes reflect some of the distinctions and conceptualizations described in the literature. However, my primary concern was accurately reflecting the breadth and diversity of experiences across the case studies, and the consistencies in challenges I faced as an evaluator working in these contexts.

The themes regarding the context of devolution consisted of, 1) accountability, 2) capacity, 3) mandate drift, 4) competition, and 5) complexity. Each theme has numerous sub-themes, which overlap and offer a variety of points of convergence between the broader themes. Table A1 in Appendix A presents a listing of the case studies from the devolution literature used to help frame and contrast the cases from my practice, and to identify and develop the broad themes and sub-themes.

I also extracted four broad evaluation themes from the cases, again using the evaluation literature and recent case studies of evaluation in local community contexts to help frame the issues and themes. While these articles and studies examined aspects of the devolution context, in most cases they did not articulate or identify them as such. The articles discussed one or more aspects of community context that also coincided with particular aspects of devolution, as represented in the first set of themes. The evaluation themes consist
of, 1) theory-based evaluation, 2) inclusiveness, 3) roles of the evaluator, and 4) dialogue and deliberation as responses to the devolution context. Once again, the themes contain a variety of sub-themes, which are discussed in detail in Chapter Six.

To date, there have not been any comprehensive treatments of evaluation in nonprofit organizations that focus on the context of devolution or the ‘marketization’ of the sector, particularly for smaller community level nonprofit or grassroots organizations. Little has been done that addresses the evolving relationships between such organizations and either government as funder, or the broader society. Through the development of these nine themes, I describe, explore and analyze the intersection between these two worlds and bodies of literature, and provide a descriptive and analytical examination of devolved evaluation in practice and impact.

**Power and Reflexivity**

I have been drawn to the literature on democratic evaluation because it addresses issues of power within the evaluation context. I have experienced this power from several angles, and as the evaluator on behalf of funders and clients have benefited from it at times. Yet in my practice I have always been aware of the element of negotiation as part of evaluation practice – negotiating the scope of the evaluation, resources and time-frame, control over methods and research approach, identifying ‘legitimate’ stakeholders, and defining the scope of the program, including its goals and activities. I have also been on the losing end of power struggles, and in other situations found myself a helpless bystander as program stakeholders were dealt with in what I viewed was an unfair or heavy-handed manner. The issue of power is always present in any evaluation study, and this calls for a measure of reflexivity concerning the evaluator’s roles and responsibilities (Gergen &
Gergen, 2002). I have found that deliberative democratic evaluation provides a useful frame for understanding and developing an orientation towards power within the research setting, and in particular a way to help reflect on the range of perspectives among stakeholders.

In the context of the present reflection on practice, my interest is not in demonstrating model examples of deliberation or democratic evaluation. Indeed, there have been times when I have stumbled on strategies that have reflected such approaches, and found them useful. There were far more occasions when a democratic orientation would have facilitated different or more satisfactory processes or outcomes of the evaluations undertaken, but in the moment I did not recognize those opportunities. However, by far the most common situation was finding myself wishing that I could employ such approaches in an evaluation, but for various reasons that I explore and discuss in the dissertation, the context itself mitigated against democratic approaches, at the same time that they would appear to be most appropriate. In part this reflects the challenges of doing democratic evaluation in “undemocratic settings” (MacNeil, 2000), and in part the unintended consequences of the devolution context.

My own reflection on power within evaluation contexts has been influenced by Saville Kushner’s deeply personal reflection on implicit power and the social location of evaluation practice in contemporary society. In Personalizing Evaluation (2000), Kushner explores the flip side of what evaluation means to those who are subjected to its intrusions into their lives. He questions the program orientation of those who have the power to request and implement evaluation activities, and to impose consequences based on those evaluations. In particular Kushner (2000) reflects on the difficulties faced by evaluators in understanding and documenting the experiences of program participants given,
a latent source of unfairness and injustice built into the fabric of program evaluation for it tends to favour the voice of those few for whom programs are useful instruments to advance their careers and their economic power. For the majority of people implicated by or involved in a program, the concept ‘program’ is barely understood and may even be irrelevant to their lives. Or, more to the point, it represents an opportunity irrelevant to their interests and their opportunities. (pp. 10-11)

Through the reflection on practice in this dissertation, I hope to transcend the instrumental concerns implicit in most evaluation practice, and see some of the ways that ‘personalizing evaluation’ can help me to understand and address differences in power among evaluation stakeholders, particularly in a context of devolved programs and evaluation practice.

A Multiple Case Study Approach

Stake (2006) describes an approach to studying and analysing multiple case studies. This approach is not completely applicable to a reflection on practice, as it emphasizes comprehensive data collection among a variety of sites, but it provides a useful framework and starting point for examining the ‘cases’ on which my evaluation practice has been based.

Multiple Case Studies

For Stake (2006, p. 6), in a multiple case study analysis, the ‘quintain’ is the “object or phenomenon or condition to be studied – a target, but not a bull’s eye.” It is the common link among the set of cases, and the starting point for analysis of the phenomenon in its many different forms and manifestations. Stake uses the example of the “proverbial blind men describing an elephant,” in which the elephant would be the quintain. In cases in which a multi-site evaluation was being conducted, the quintain would typically be the program.

For Stake (2006, p. 1) the ‘case’ of a case study is “… rather special. A case is a noun, a thing, an entity; it is seldom a verb, a participle, a functioning.” A community or

---

2 This work builds on and extends ideas from Stake’s classic The art of case study research (1995), in which he refers to multiple cases as ‘collective case studies.’
agency can be a case, as could something as amorphous as a program or a training module, but not generically ‘services to clients’ or ‘training’ in themselves. We examine cases in which there are “… opportunities to examine functioning, but the functioning is not the case” (p. 2). The cases we examine are bounded, although they may also be complex, situational, and connected to diverse issues, interests and even other cases.

While it might be tempting to think of a case as a ‘snap-shot’ picture of a program or situation, Stake (p. 3) emphasizes viewing cases as dynamic and operating in real time, and as embedded in historical contexts that involve many individuals and events, each of whom could be a case. In this way virtually everything about a case becomes potentially relevant to studying the quintain – history, political, cultural and social contexts, activities, documents, and other efforts to examine the case, which could encompass other evaluations, previous programs, and the broader programming environment in which the organizations operate.

For this study, the quintain is the practice of evaluation, and the sites are the specific nonprofit contexts of devolved program delivery with which I have worked over the past dozen years – the communities, agencies and programs in which the evaluation work took place. The evaluations on which I draw are diverse, from small and short-term studies encompassing a few months of activity, to complex multi-site studies consisting of up to ten individual case study communities, and with an implementation of the evaluation lasting up to three years. The communities and program contexts that represent the cases in which the evaluations took place also represent extreme diversity, ranging from large urban centres (Vancouver, Victoria, Surrey) to small rural communities of a few hundred people.

The purposes of case study research can be either ‘instrumental’ – going beyond the case, or ‘intrinsic’ – looking at the key elements of the case that are of enduring interest in
and of themselves (Stake, 2006, p. 8; Creswell, 1998, p. 62). For multiple case study analysis, the emphasis is usually on the instrumental case – and this study is no exception. While the specific and unique attributes of cases are relevant to my analysis, the focus is on understanding evaluation practice across the cases, and seeing the larger picture of which any one case might contribute one or two pieces of the puzzle.

Data for the Case Studies

As a reflection on practice, the study examines details from cases that spanned the past dozen years of my practice. Many of the people I worked with over this period – clients, program clients, program and service deliverers and administrators – are no longer associated with the organizations in which they were involved at that time, and indeed, many of the programs, projects and organizations no longer exist in the form in which I experienced them. A retrospective examination of cases is challenging but possible; a retrospective collection of information from evaluation work conducted in the past is not achievable.

The case studies use existing documents and public records, my own notes and records from the evaluations, and reflective materials that represent background information used in developing and negotiating contracts and draft reports. I also use several reflective narratives describing my own efforts at grappling with devolved evaluation issues, constructed as part of my course work for the doctoral program.

Stake (2006, p. 23) notes the main criteria for selecting cases:

- Relevance to the quintain,
- Diversity across contexts, and
- Cases that provide good opportunities to learn about complexity and context.

While not all of the cases from my practice over the past dozen years have equal relevance for understanding the devolution context, most have some relevance. As a group,
they represent a very diverse set of contexts, although certainly not exhaustive of all possible situations relevant to understanding the quintain for my study. Most do provide opportunities for learning about evaluation under devolution, and they are particularly appropriate for addressing the complexity of devolved evaluation in local communities.

**Triangulation**

One of the reasons for conducting a multiple case study is that it provides opportunities to enhance validity – in looking at multiple cases, the researcher focuses on factors that cut across several cases, and appreciates the interplay of special circumstances in contrast with those that appear to me more universal. Examining cases retrospectively – and despite how disparate they might be, all from one evaluation practice – represents a challenge to the study’s validity. I am reflecting on my own practice, and although I have frequently discussed devolution and its impact on evaluation with other consultants and evaluators, I have not ‘interrogated’ their practices to the same depth as I have examined my own.

Yet the evaluation discipline, as well as the practice-based literatures of program delivery and administration have made excellent use of case studies as means to share understanding about the context of practice. Indeed, examining these case studies has been a crucial element of my efforts to understand and develop the concepts used to make sense of my practice. I have reviewed a wide range of case studies reflecting either 1) devolution of programs and services, or 2) evaluation in community-based nonprofit organizations. Jensen and Rodgers (2001) argue that meta-analysis of case studies allows researchers to resolve problems of knowledge cumulation and generalizability. Such meta-analysis represents a form of data triangulation as discussed by Denzin (1978) and Patton (2002a).
A few of the evaluation case studies explicitly examine issues relevant to the devolution of programs and services as a context for the evaluation, but that is rare. Very few of either set of case studies address grassroots organizations specifically. Most case studies are bounded much more narrowly than the group of cases from my own practice. A few cover a fairly wide time span, such as five years (Clayson, 2002), but usually within a fairly bounded community and geographic range. These case studies provide invaluable reference points for framing and contrasting the examples from my own practice. In many cases the findings of these studies complement the observations from my own practice. In some they are at odds with my experience and interpretations, and this helps point to areas in need of further research and clarification.

**Composite Scenarios**

Because a key element of my reflection on practice is looking at the complexity of contexts in which evaluation is practiced, and also because of the diversity of the contexts I link through my analysis, I present the evaluation case studies as a series of vignettes representing two broad ‘composite scenarios,’ one describing a grassroots organization, and the other describing a small nonprofit organization. This serves several purposes for the study and for the dissertation: it addresses the complexity and breadth of the cases and makes them more comprehensible to the reader, and it addresses confidentiality concerns posed by a ‘reflection on practice.’

Stake (1995, pp. 86-87) discusses how the detailed descriptions included in case studies can assist with ‘naturalistic generalization’ – learning much that is general from individual cases, trying to understand the implications of those cases in other contexts. Because the cases I describe are so diverse and the contexts so rich and varied, discussing
each of them would be quickly overwhelming to the reader, and presenting the volume of
contextual detail would represent a daunting challenge. Further, the links among cases and
the broader themes are facilitated by such a composite approach – the factors are not
independent of one another, despite the fact that no one case would provide avenues for
seeing every ‘part of the elephant.’ Stake (1995) suggests that,

To assist the reader in making naturalistic generalizations, case researchers need to
provide opportunity for vicarious experience. Our accounts need to be personal,
describing the things of our sensory experiences, not failing to attend to the matters
that personal curiosity dictates. A narrative account, a story, a chronological
presentation, personalistic description, emphasis on time and place provide rich
ingredients for vicarious experience. (p. 86)

Yet to be comprehensible, the detailed accounts benefit from being consistent in presentation,
shorter and more concise in exposition, and because of the number of issues to be discussed,
built on continuity and connection. The presentation of the cases using two broad scenarios is
part of the process for providing an analytic frame for understanding the relevance of what is
being discussed, and doing so in a way that simplifies the narrative arc. Kearns (2003)
provides an example of such a composite scenario process, used for illustrative purposes as a
way of clearly portraying a complex context. At the end of this chapter are two background
vignettes that map out the parameters of the composite scenarios. In creating these scenarios,
I have conceptualized them as two points of a continuum that would range from very small
grassroots organizations to large, multi-site nonprofit organizations or foundations. One
scenario represents the grassroots end, and the other a small nonprofit organization that is
still close to that end of the broad continuum, but closer to the medium or larger nonprofits in
characteristics and broadening program focus.

The composite scenarios are further portrayed in Chapters Five and Six through the
use of another twenty-three detailed vignettes, each facilitating reflection on a particular
theme or sub-theme. Most of these vignettes can be viewed as excerpts from the multiple cases that represent the data for the study. The individuals portrayed in the vignettes are composites, using standardized names, characteristics and roles to support the clarity of the presentation and also to safeguard the identities of community members. Several of the vignettes also represent composites, in that rather than simply portraying an event or series of events from the cases, multiple settings and in at least one instance, several evaluation studies have been compiled to form one clear vignette to demonstrate the point in question.

Another part of the rationale for composite scenarios is the issue of confidentiality. As a retrospective approach, the reflection on practice examines issues and settings for which there exist no opportunities for obtaining permission and consent. The evaluations themselves, and in many cases the reports prepared for them, are public processes, and part of the natural process of my work, and the products are a part of the public record. I did not conduct interviews or focus groups related to my dissertation research as part of doing this work. However, in the process of conducting interviews and focus groups for the research projects, relevant topics arose and became part of my understanding of the program or setting being examined. Insofar as the interviews I conducted were used in constructing reports, I can use that information as a public representation of issues that is also relevant to this dissertation. Yet those conversations also contained information that was not intended by the subjects to be included as part of the evaluation studies, and in many cases this information was more relevant and interesting for my understanding of the context of the study, rather than for the study itself.

Further, in grappling with and developing these ideas over the past four years – and indeed, the past decade – I have gone through an intensive but informal process of discussion
and reflection with clients, colleagues and project informants. Many of the conversations that informed my development of the ideas and understanding of the dynamics of the changes occurring in nonprofit programs, agencies and accountability relationships that I discuss in this dissertation have been ‘in the margins.’ These conversations have occurred as ‘small talk’ at the beginning and end of interviews and focus group discussions as informants reflected on their organizations and on their own participation in ‘evaluation’ of programs. They have occurred in the negotiation of contracts and identification of roles, responsibilities, priorities and constraints. They have occurred in the hallways between sessions at conferences, in the coffee breaks of workshops and training sessions I have attended and delivered, and in informal conversations with those developing government policy and programs. In short, they have occurred in the margins of my ‘official’ working life – not as parts of my contracts or evaluation work, nor in my official capacities as researcher or graduate student. They helped me to understand and appreciate devolved evaluation, but they are not appropriate for direct quotation or inclusion in case studies drawn from specific cases.

Even so, the people I have spoken with have been passionate about the changes affecting them, and they have often jumped at the opportunity to talk about issues central to their working lives, over which they find few opportunities to either reflect or take action. In my informal discussions with them, most individuals expressed interest in my observations, and were quick to offer suggestions of examples of themes and their own concerns. Such discussions were not systematic data-gathering, but they did point me towards examples that existed within the cases I was already examining, and represented ‘found’ opportunities to better understand and articulate the nature of the devolution of programs and services. These
conversations typically informed my understanding of the contexts of practice for those delivering programs and services, rather than the evaluation of those programs.

**Composite Scenario Background**

In order to contrast the various types of case I have examined in my practice, and to better contrast the particular challenges of grassroots evaluation from that in larger nonprofit organizations I have developed two composite scenarios from the multiple cases on which I am drawing. One is of a recently incorporated grassroots organization that has grown through the efforts of a group of parents dealing with the needs of their children. The other is a small, established nonprofit organization that has been delivering programs and services in one community for approximately twenty years, and although not yet affiliated with a larger umbrella organization of deliverers, is in a position to contemplate doing so.

**Vignette 2.1 – The Grassroots Composite Scenario**

Laura is the director of a small grassroots organization in a community with a total population of less than ten thousand people. The organization grew out of the efforts of a group of parents and teachers seeking to meet the needs of children with developmental disabilities. Parents and caregivers are finding it difficult to obtain recognition of their challenges, and are also finding few supports within the school system or afterwards in the community as they try to enter public life in the community. The organization has a board composed of parents, teachers and some service providers, has recently gone through the process of incorporating, and the board is examining the benefits and requirements of obtaining official ‘charitable’ status.

Over the past ten years the organization has received seed money from provincial and federal departments and agencies to develop a variety of public education programs and
services within the community. The education is targeted broadly – to schools, parents, broader community members, and those within the program and service delivery network – other agencies and government employees with whom the parents and their children of necessity interact. The education sessions provide information about the specific form of disability in question, and how it is relevant to the needs and capabilities of the young people, as well as the implications for service delivery to this population. The group formed on the basis of a perceived need to advocate for their children locally, provincially and nationally, and moved into public education as a means of addressing the needs of young people more broadly within the immediate community.

The board and agency are at the point of growing again, as they have hired a director (Laura) and two staff people to work with the many volunteers who have been working with the organization over the years. All of the staff members now with the agency began their involvement as volunteers with various initiatives, and indeed many still do volunteer part of their time as well as working for the organization. The organization coordinator and board members have developed several innovative approaches to programming, which they began offering in the community using volunteers, and presented at conferences of service providers. Based on these conference presentations, they have obtained funding from federal and provincial agencies to develop their innovative approach, which fills gaps in services within the community and outlying area.

The new funding stipulates that the project must evaluate its efforts, and provide ‘lessons learned’ and ‘best practices’ that can be shared with other communities. Both sets of funding are parts of broader funding envelopes that are not delivering specific government programs, but supporting innovative community-based solutions to public issues. The
funding sources are diverse, and represent agencies and departments with education, health, youth mental health, social services and ‘First Nations’ support program focuses. The evaluation efforts are required for most of the projects, and a small amount of the funding has been budgeted for supporting evaluation. The organization board has chosen to use the evaluation requirement to examine all facets of their project, as it has grown in several directions, and so has hired a consultant to coordinate this work, and to support and document their evaluation efforts. Part of the rationale for this evaluation is to enhance the organization’s ability to position itself with respect to future anticipated growth, and possible linkages with other similar organizations that are forming across the province.

As the evaluation consultant to the organization, I have been asked to work with the director, staff and volunteers in a participatory approach. Part of the rationale for this is that they cannot afford to have an evaluator complete all of the work – their budget is small. But a participatory approach also fits the organization’s rationale and orientation towards its role in the community, which emphasizes a community development focus based on developing relationships with and among diverse stakeholders, agencies, and front-line deliverers.

**Vignette 2.2 – The Small Nonprofit Composite Scenario**

Susan is the executive director of a small nonprofit in a northern community of 25,000 people. The agency has been active in the community for approximately twenty years, and has offered a variety of community health, mental health, health education and health promotion programs over that period. The agency has a dozen full and part-time employees, and regularly has up to forty volunteers participating in program activities in the course of a year. The organization has built up strong links with other community agencies and organizations, and membership on the boards of these agencies sometimes overlaps.
Looking to stabilize the funding for community programming, Susan has recently been successful in expanding the agency’s programming after submitting proposals to provide local delivery of several provincially and federally funded programs that had not yet been established in the community. Two of these programs involve connecting with a larger network of deliverers in other communities, working through an umbrella organization that mediates between the funder and community delivery partners. Another venture is a demonstration project active in eight communities across the province, and part of a two-year initiative that is examining an innovative approach to community delivery. In order to deliver these programs, Susan has rationalized the delivery of several services that had traditionally been delivered by the agency, but funded through municipal and community fundraising efforts, and primarily delivered by volunteers.

I have been working with the demonstration project to evaluate the innovative effort in eight communities, and have been using a participatory approach through which I facilitate evaluation planning and data collection in the communities, but much of the work is done locally. I have arranged to visit each of the communities twice per year, and have their local lead contact visit Vancouver twice per year as well, to share stories and strategies for the project evaluation efforts of each of the eight communities. Susan has been my initial contact for the demonstration project evaluation, as she wrote the proposal and oversaw the development of the initiative. After attending the first gathering in Vancouver, Susan has since passed on the responsibility to the project coordinator Sally, who has been hired specifically for the project, and may or may not continue with the agency beyond its completion. The project has been established with base funding that cannot cover all of the project expenses; funding is offered with the stipulation that the agency and local community
also contribute matching funding or resources representing at least a third of total funding. None of the funding offered through the project may be used for capital expenses, and evaluation of the project is a required element.

My efforts to develop and support the evaluation of the project demand that I spend time with Susan discussing the information systems used for all of the agency’s programming, and not just the one project we are jointly working on. Susan has been working with provincial and two federally funded programs to develop an information system that collects basic information about participation in programs and services, and at the minimum provides output data related to core agency functions. Susan and the board are supportive of this effort, as they would like to have a better sense of what is happening with the agency as it grows. The agency’s programs and services are currently delivered at two locations within the town, and through two other locations in nearby First Nations communities. The geographic expansion has challenged Susan’s ability to keep track of what is being done in the other offices, and she hopes that the information system being developed represents an opportunity to reduce her travel demands to other offices.

The project that I am working with is being delivered through three of the four community locations, and I have made an effort to visit those offices to help put a face to the name being mentioned as ‘the evaluator.’ Most staff and many volunteers met with me on my first visit to the community as the project was starting up, and support across the organization seems high and enthusiastic. It is a time of growth in the organization, and the community appears to support the initiative, and has expectations of success and further expansion, which they are eager to document.
CHAPTER THREE: TRENDS IN THE NONPROFIT, VOLUNTARY AND GRASSROOTS SECTOR

The conventional designations of the nonprofit sector do not give us the tools we need to conceive the question. Here is why: the nonprofit sector is not really a sector, but rather a residual category. It consists of organizations that are neither government nor for-profit business. The residual nature of the term plagues similar designations: the concepts of an independent sector, a third sector, a voluntary sector, and civil society, are no more helpful.

Mark E. Warren, 2003, p. 47.

This chapter summarizes and reflects on the literature examining the status of nonprofit organizations in Canada. The two main sections of the chapter examine the current profile of nonprofit organizations, and some of the history of devolution of programs and services to nonprofit organizations, respectively. Because many of the examples and case studies in the literature reflect the nonprofit context in Europe, Southeast Asia, and the United States, I will also discuss the nonprofit in an international context, and reflect on some relevant broad global trends, including recent discussions on civil society.

Across North America, nonprofit, voluntary and grassroots organizations are in transition. Indeed, the devolution of programs and services from all levels of government to third sector organizations has been a worldwide trend over the past two decades (Day & Devlin, 1997; Alexander, 2000 & 1999; Rekart, 1997 & 1993; Nowland-Foreman, 1998; Salamon, 1993). This trend has intensified since the early 1990s, reflecting neo-liberal political changes, as well as intensified competitive pressures associated with globalization (Richmond & Shields, 2004, p. 54). While the trends have produced significant and widespread impacts on the operations of nongovernmental organizations (NGOs) – particularly nonprofit, voluntary and grassroots organizations – there has been a curious gap in analysis and writing about these issues, not just in Canada, but also around the world (Evans & Shields, 2005; Richmond & Shields, 2004).
Defining the Nonprofit Sector

There is not a great deal of consistency in the literature regarding either the boundaries of the nonprofit sector, or in the terminology used to describe it. This applies to distinguishing the sector from the public and private spheres, as well as identifying relevant features of the nonprofit sector itself. Dreessen (2000, p. 2) attributes this to the diversity of purposes for analysis and classification to which conceptualizations of the sector have been subjected. Similarly, Hirshhorn (1997) attributes the different terminology used in describing the sector to the diversity of disciplinary approaches to the subject matter:

… different labels – nonprofit, not-for-profit, voluntary, third or independent sector, “The Commons” – have their roots in different scholarly approaches to thinking about the sector. While, for example, sociologists have been attracted by the noncoercive aspects of participation in the “voluntary sector,” political scientists have focused on the potential benefits to pluralist democracies from the availability of an “independent” or “third sector.” By adopting the term “nonprofit” throughout the project, we have subscribed to the language of economists. (p. 5)

Given the diversity of organizations that are considered to be part of the nonprofit sector in Canada and globally, the diversity in terminology in use is not surprising. As a ‘residual category’ (as noted by Warren, 2003, p. 47), the nonprofit sector includes everything from hospitals and universities to arts organizations, social clubs, day-care centres, advocacy organizations, unions, places of worship and organizations that deliver community level services and programs. The organizations can vary in size from a few individuals to thousands of paid staff and volunteers.

Nonprofit organizations and the nonprofit ‘sector’ are consistently called by such names as ‘nonprofit,’ ‘non-profit,’ ‘not-for-profit,’ ‘non-governmental organization’ (NGO), ‘community-based,’ ‘charitable,’ ‘voluntary,’ ‘independent,’ ‘third sector,’ and sometimes the whole sector is referred to as ‘civil society.’ What they typically have in common is that they are institutionally separate from government, and are not commercial businesses. They
may obtain revenues from sales of products or services, but do not distribute ‘profits’ to owners or directors. According to Statistics Canada (2004), they are self-governing (independent and self-regulating), and voluntary (using volunteers or donations of time or resources). While these features help us to define the sector, they are not conclusive.

Baulderstone (2005) calls the nonprofit sector a ‘fuzzy’ sector, with blurred distinctions between public, private and third sector organizations. Usually we think that:

… private sector organizations undertake business and commercial activity, public sector organizations provide physical infrastructure and deliver services to the public, and nonprofits reflect special interests. In practice, third sector organizations may engage in commercial activities and make a profit, they may compete with private sector organizations for work, and governments contract with such agencies to provide services on their behalf. Government agencies and some private sector organizations use volunteers and both government and private sector organizations support nonprofit organizations in various ways. (p. 4)

In practical terms, the boundaries between the sectors overlap when we look at individual criteria, yet we usually do not have difficulty determining the sector of specific organizations. Some of the arguments concerning the changes to the nonprofit sector have focused on the ways in which the sector is becoming more like both the private and government sectors – the ‘marketization’ of the sector and the ‘hollowing out’ of government, respectively. I will return to these themes in the next section.

For the purposes of this dissertation I will use the terms ‘nonprofit organization’ and ‘nonprofit sector’, except where quoting other writers, or referring to specific sub-groups within the broader sector.

Distinguishing between types of nonprofit organization remains the most significant challenge in defining the sector. Recently, Salamon and Anheier (1997) developed international criteria for classifying nonprofit organizations, and in Canada, this provided the framework for Statistics Canada’s recent (2004) studies on nonprofit organizations. Even so,
the distinctions that this classification scheme makes do not adequately address the need to describe and analyze the varieties of organizations within the sector.

In the United States, classifying nonprofit organizations appropriate for this study would likely be an easier task, as there is a simple tax-code designation for such organizations. The category of 501 (c)(3) covers charitable organizations, and 501 (c)(4) describes organizations with social welfare purposes (Warren, 2003, p. 47). Similarly, in Australia, in order for charitable organizations to obtain government funding, or to compete for service delivery contracts, they must be legally incorporated (Baulderstone, 2005, p. 5). There is no comparable requirement in Canada, and at their discretion, federal or provincial departments or agencies can and do contract with unincorporated ‘grassroots’ organizations.

**Canadian Nonprofit Organizations**

Within Canada, several recent national level studies and initiatives have begun to document the sweeping transition of the nonprofit sector. Statistics Canada’s study Cornerstones of community: Highlights of the national survey of nonprofit and voluntary organizations (2004) identified fifteen categories of nonprofit organization, based on their primary activity areas. For this study, the main types of nonprofit organization with which I worked and that provide community level delivery or programs and services would be found in the ‘Social Services’ category, comprised of “organizations and institutions providing social services to a community or target population.” The 19,099 organizations identified through Statistics Canada’s study represented 11.8% of all such organizations. And yet, on closer examination, some of the nonprofit organizations that I have worked with could easily be included under the categories of ‘Health,’ ‘Education and Research,’ ‘Development and Housing,’ or ‘Law, Advocacy and Politics.’ Further, although the nonprofit organizations are
categorized based on their primary activity, with the expansion and contraction of
organizations based on the number and type of program delivery contracts that they take on,
nonprofit organizations can have substantial transformations of their range of activities, key
programming areas, and how they define what they do. Further, the categories are not
mutually exclusive, as organizations that, for example, provide health education to teens
might easily choose ‘Education and Research,’ ‘Health,’ ‘Social Services,’ or ‘Development
and Housing’ as the best-fitting category, depending on how they were thinking about the
activity as either a health promotion, education, community development, or program
delivery activity. Clearly it could be considered all of these at once, but which would get the
nod as securing the identity of the agency? And which program among perhaps a dozen that
might be offered by the organization? The categorization would also reflect who in the
organization participated in the survey, as a program manager or front-line staff person might
have a very different idea of the organization’s activities than would an executive director or
board member. See the table in Appendix B for a portrayal of the complex picture of
nonprofit organizations in Canada.

Perhaps a larger issue for this study is that Statistics Canada’s National Survey of
Nonprofit and Voluntary Organizations (NSNVO) excluded grassroots organizations and any
groups that were not formally incorporated or registered with provincial, territorial or federal
governments (2004, p. 7). This reflects the difficulty in identifying and locating the
organizations. Dreessen (2000) notes that with regard to unregistered nonprofit
organizations:

… two conclusions are inescapable: one, noncharitable nonprofits are by no
measure a negligible phenomenon; and two, we are almost completely in the dark
about even the most basic data for this segment of the sector. (p. 13)
Similarly, Day and Devlin (1997, p. xx) point out “… it has proven to be almost impossible to obtain an accurate measure of the size of the noncharitable nonprofit component of the sector.” Indeed, Smith (1997) asks why grassroots associations are so seldom studied by those studying the nonprofit sector. He suggests that perhaps it is because they look at individual organizations using a bureaucratic model rather than using a more appropriate market model, which would emphasize the system of groups in the community in competition with one another for resources, or linked together within a network of program and service delivery. This perspective fits with an observation that grassroots organizations are hard to find – unless you ask people in other community organizations where they are.

The Voluntary Sector Initiative (VSI), a five-year joint initiative between the government of Canada (providing most of the financial support) and a variety of third sector organizations – particularly the Canadian Council on Social Development – was active between June of 2000 and mid-2005 in conducting research and preparing reports on the status of nonprofit organizations in Canada. Statistics Canada’s survey (2004) reported that Canada had over 161,000 incorporated nonprofit and voluntary organizations in 2003. The study reported findings from a survey of 13,000 of these organizations, which were identified from a database of incorporated nonprofit and charity organizations. If we were to include grassroots organizations, Evans & Shields (2005, p. 2) and Hirshhorn (1997, p. 8) estimate that the number of organizations would likely exceed 200,000. The NSNVO included a diverse range of organizations, from large universities and hospitals, to small voluntary associations and agencies with annual revenues of under $30,000. Just over half (56%) were registered charitable organizations.
Evans and Shields (2005, p. 3; & 2002, p. 140) divide third sector organizations into four categories: 1) funding agencies such as the United Way, 2) public benefit organizations such as day cares and social service agencies that provide goods and services – programs for the general welfare, 3) member serving organizations such as trade unions, business or professional organizations that serve a membership group rather than the public at large, and 4) religious organizations such as churches involved primarily in sacred and religious activities. This approach is similar to the nonprofit groupings used prior to recent attempts to develop more complete databases about the sector. They offer a simpler categorization than the Statistics Canada survey, but it also excludes a wide range of larger nonprofit organizations that are part of the diversity of the sector – hospitals and universities.

Some of the key findings of the Statistics Canada survey relevant to this dissertation include observations that larger organizations receive the bulk of government resources flowing to nonprofits, are more dependent on government funding, and the largest organizations are also those that are growing – they are more likely to have reported increased revenues, more volunteers or a higher number of paid staff over the previous four years (between 2000 and 2003). By contrast, smaller organizations were more likely to report declining revenues, fewer volunteers and stable staff levels for the same period (Statistics Canada, 2004, p. 10). The majority of organizations reported challenges in fulfilling their missions based on capacity – difficulties in recruiting and retaining volunteers and board members, and difficulty obtaining funding.

In short, the report described a nonprofit sector in transition, experiencing significant challenges and stresses, yet representing an enormous presence in the Canadian economy, with revenues in excess of $112 billion, $75 billion of which is not in hospitals, universities...
or colleges. Nonprofit and voluntary organizations employ over two million people, although hospitals, universities and colleges employ a third of this number, but make up less than 1% of all organizations, and use over nineteen million volunteers who contribute over two billion hours of volunteer time – representing the equivalent of a million full-time jobs per year. Just under half (49%) of the funding for the sector originates directly in government grants and contracts.

**Civil Society**

As Warren (2003) notes, one of the more interesting ways to understand the nonprofit sector is as ‘civil society,’ which has been a focus of many of those writing about devolution (Alexander, 1999 & 2000; Eikenberry & Kluver, 2004; Brock, 2002; Evans & Shields, 2005; Nowland-Foreman, 1998; Wolch, 1999). Alexander (1999, p. 454) refers to civil society as “the space occupied by associations,” and this seems to represent the starting point for how civil society has been used in this literature. Most say they are building their conceptions of civil society on de Tocqueville’s work of over a century ago, and for these writers, the idea represents a core feature of workable democracy in a capitalist society – a structural element of society representing that point of overlap and convergence between the state and the private sector; part of neither, but mediating the effects of both. For many writers such as Rifkin (1995), civil society represents the next realm of opportunity in the face of technological impacts – and a source of hope.

Edwards (2005) takes issue with the lack of clarity and rigour among those using the term ‘civil society,’ arguing that it is used in so many ways that it is in danger of becoming meaningless. He argues that if we look at civil society as a ‘part’ of society – nonprofit or voluntary organizations – then we lose the original and still useful meaning of the concept –
civil society as a kind of society identified with the ideals of political equality and peaceful coexistence.

Edwards contrasts three ways of looking at civil society: 1) as associational life, 2) as the good society, and 3) as the public sphere. Civil society as associational life emphasizes a structural approach, viewing the nonprofit sector as the context in which the good society develops. Like Putnam (2000), Edwards notes the danger of this perspective – that nonprofit organizations are home to competing and anti-democratic values as well as those that support democracy. He also notes that in terms of expecting this sector to be a saviour, we actually spend a small part of our total lives in such associations, and it is their relationship to families, schools and other institutions that the nonprofit organizations have their key impacts on society.

Civil society as the good society, as the idea of what kind of society we want to live in, always leads to the question of how do we decide what is good? Civil society as public sphere – “the places where citizens argue with one-another about the great questions of the day and negotiate a constantly-evolving sense of the ‘common’ or ‘public’ interest,” (Edwards, 2005) provides a process for getting there – through having a public space in which we can deliberate, negotiate, and work out our definitions of problems and questions, and our solutions to both. Edwards argues that civil society remains a useful concept when we think of it as the arena in which we can do this kind of deliberation. The more structural definitions of civil society help us to identify “gaps and weaknesses in associational life,” (Edwards, 2005) but de-emphasize the relationships between sectors, and promote a view of the nonprofit sector that overemphasizes its role and capability.
This view of civil society resonates with the work of Lyons, who contrasts the nonprofit sector and civil society. Lyons (1996, pp. 3-6) notes that the literature on the ‘nonprofit sector’ has grown out of work by economists, who have emphasized quite structural definitions of the sector based on the American tax-status distinctions noted earlier. This has led to the exclusion of some types of third sector organizations, notably mutual aid organizations, cooperatives and self-help groups – types of association that might inhabit the grassroots sector: nonprofit organizations in the early stages of becoming part of the broader social service delivery web. In contrast, Lyons notes that ideas about civil society have grown out of the sociological literature, encompassing conceptual rather than concrete and pragmatic concerns, and have focused specifically on relationships among people and organizations.

For example, the nonprofit literature views volunteering as philanthropy – self-interested giving as part of a reciprocal and symmetrical exchange – while civil society views volunteering as membership and as a form of collective action that can contribute to the development of social capital (Lyons, 1996, p. 12). While much of the economics-based nonprofit literature seems comfortable with viewing nonprofit organizations as a special form of firm, using corporate models of governance, and addressing the needs of consumers as ‘customers,’ the civil society literature focuses on democratic forms of governance, and the capacity of the sector to encourage participation in social life – viewing participants as citizens rather than strictly as consumers of services (Lyons, 1996, pp. 12-13). In essence, the two literatures refer to different organizations and types of organization, start from differing assumptions about the form of the sector, and reflect potentially conflicting views of the roles, activities and purposes of the organizations they encompass.
The most common interpretations of civil society in the devolution literature remain fairly structural – reflecting their overlap with broader literature on the ‘nonprofit’ sector. In view of the diversity of perspectives on civil society, and the lack of conceptual clarity among many using the term embedded within broader discussions growing out of the literature on nonprofit organizations, for the purposes of this dissertation I will continue to use the term ‘nonprofit’ when discussing the sector, unless quoting a specific author.

**The Devolution of Programs and Services**

While solid statistical information on the history of growth in the nonprofit sector in Canada is quite limited, it is clear that the 1980s and 1990s were a time of significant growth. Rekart (1997, p. 3) notes that both the federal and provincial governments became increasingly involved in providing services directly during the middle decades of the last century, in effect taking over and professionalizing the delivery of the welfare state’s safety net from charitable organizations – particularly religious organizations – and municipal governments that had undertaken such responsibilities in the recent past. The evolving relationships between the federal and provincial governments have provided a constantly shifting and negotiated backdrop to the delivery of programs and services, with the long-term trend being movement to government initially, and then a gradual shift from the federal government to provincial government jurisdictions, and finally devolution back to the nonprofit and voluntary sector over the past two decades (see Hall & Reed, 1998). For the latter devolution, this involved concurrent shifts in funding by government, as well as control over budgets and products.
The Rationale for Devolution

The growing literature about the impacts of nonprofit sector restructuring remains diverse and somewhat disjointed, although there is a growing body of research on the process and impacts of devolution. Theoretical work has tended to look at the origins of restructuring in terms of the neoliberal foundations that are transforming government, private sector and third sector financing. The work has also begun to explore the transitions in governance forms that have been associated with devolved program and service mandates (Evans & Shields, 2005; Eikenberry & Kluver, 2004; Richmond & Shields, 2004; Brock, 2000 & 2002; Alexander, 1999 & 2000; Rekart, 1993 & 1997; Nowland-Foreman, 1998; Salamon, 1989, 1993, 1997 & 2003; Pulkingham, 1989; Wolch, 1999).

On a practical front, the issues of fiscal reform and the impacts of devolution on nonprofit organizations and their clients has been a hot topic in conferences and gatherings of those working on the front lines. The specific impacts are commonly discussed – but often in isolation, and often with little connection to the broader societal trends in which they are embedded (Richmond & Shields, 2004; Phillips & Levasseur, 2004) or to other stakeholders who might have useful ideas or observations to contribute. As such, there exists a clear need for a more comprehensive and interdisciplinary treatment of devolution in the Canadian context, and one that can bring together the various strands of the current debate in a meaningful and useful way for those affected. Although this remains a worthwhile goal, it is beyond the means of this dissertation to provide such a comprehensive examination of devolution. Even so, as a precursor to understanding devolved evaluation, in this dissertation I do provide a framework for conceptualizing the impacts of devolution, particularly focusing on grassroots organizations.
In the United States, much of the literature concerning devolution has grown out of welfare reform, and the shift of delivery of welfare out of government. This shift has involved two facets that are not present in the Canadian context. The first is that the delivery of welfare programming itself has not been devolved in Canada. While many ancillary and complementary programs and services, such as employment training and support services, have been devolved – inconsistently across the different provinces – for the most part the administration of the welfare system is still undertaken by civil servants. The second facet of devolution that differs between the two countries is the level of introduction of the private sphere into program delivery. In the United States the marketization of welfare and other programs has introduced direct competition between nonprofit and private sector deliverers. While there has been some growth of private sector organizations in Canada delivering some programs, and certainly in the private school market, the direct competition between the sectors has remained more a ‘myth’ among nonprofit practitioners rather than a practice or strategy employed by government funders. Indeed the Canadian nonprofit sector has been watching the process of devolution in the United States, and in recent years conversations have been rife with portents of future private sector competition for program delivery.

Lester Salamon (1989, 1993, 1997 & 2003) has been at the forefront of writing and researching the devolution trend in the United States for the past twenty years. In talking about a theory of government-nonprofit relations in the welfare state Salamon (1989) describes some of the practical advantages that make the nonprofit sector inviting as a deliverer of human services. These advantages include,

- a significant degree of flexibility resulting from the relative ease with which agencies form and disband and the closeness of governing boards to the field of action.
• Existing institutional structures in a number of program areas because voluntary
  agencies often begin work in an area prior to government getting involved in these
  areas.
• A generally smaller scale of operation, providing greater opportunity for tailoring
  services to client needs.
• A greater capacity to avoid fragmented approaches and to concentrate on the full
  range of needs that families or individuals face, to treat the person or the family
  instead of the isolated problem.
• Greater access to private charitable resources and volunteer labour, which can
  enhance the quality of service provided and leverage public dollars.

Salamon (1993) discusses the marketization of welfare and how it is changing
nonprofit and for-profit roles in American program delivery. He defines marketization as
“the penetration of the essentially market-type relationships into the social welfare arena”
(1993, p. 17), and notes that as for-profit firms enter the social market, they will inevitably
siphon off the more affluent customers, leaving nonprofit firms with the most difficult, and
least profitable, cases.

In general, the stated rationale for the devolution of programs and services to the
nonprofit and voluntary sector under the New Public Management has included reducing the
costs of public services, improving efficiency, and reducing the size of the state (Evans &
Shields, 2005; Eikenberry & Kluver, 2004; Alexander, 1999 & 2000; Brock, 2000; Rekart,
1993 & 1997; Trebilcock, 1995; Ferris, 1993; Salamon, 1993; and Pulkingham, 1989). Some
of the more recent discussions also identify such factors as finding ways to bypass public
sector unions, and managing results with a customer service orientation (Phillips &
Levasseur, 2004). The emphasis in introducing competition in the delivery of programs and
services – marketization – was particularly seen as a means of introducing efficiencies, and
this has also been the rationale for privatizing programs and services as well.

While much of the literature has focused explicitly on the nonprofit sector, and has
been concerned with the impacts on organizations taking on new roles and responsibilities, a
complementary literature has examined the impact of devolution on the state. Rekart (1997), Nowland-Foreman (1998) and Wolch (1999), among others have examined the ‘hollowing out’ of the state sector. Their concern is with the expansion of the nonprofit sector as ‘shadow state’ – as taking on the roles discarded by the state – and with the reduced capacity of the state to intervene in the public interest, combined with the questionable capacity of the nonprofit sector to successfully take on such responsibilities.

Miller (1998), in examining the history of devolution in Canada, notes the 1995 federal budget as a turning point for Canadian nonprofit organizations as federal transfer payments to the provinces were simultaneously restructured and reduced. Because of this, provincial funds flowing to nonprofit organizations were also cut or terminated, and began being restructured as service contracts continued to replace grants. He examines one argument that suggests the vulnerability of nonprofit organizations in this environment has grown in part from their over-dependence on government during the previous 25 years.

Burnley et al. (2005) in a study of local agencies delivering programs and services for children and families, examined impacts of devolution in Nanaimo, B.C. They note that in response to provincial and federal deficits (referencing Brock, 2000), and in the face of rising public demand for services, both levels of government proceeded to devolve delivery of many social programs to nonprofit organizations. In British Columbia, in response to an administrative review of government practices (the 1993 Korbin Commission), the provincial government initiated a contract reform project “intended to reduce fragmentation in its approach to contracting services with the social and community services sector. The stated outcome objectives of the project were to establish long-term relationships with eligible contractors, improve consistency and coordination of contracting practices, streamline
administration, and increase accountability” (p. 71). More recently, the provincial government has developed “province-wide standards for locally delivered programs, the establishment of accountability and performance management mechanisms, and the increased use of output and outcome-based contracts for service” (p. 71).

The discussion by Burnley et al. (2005) highlights a key aspect of the transition to the era of New Public Management; it has brought a variety of reforms to how programs and services are funded. Devolution involves a move to contracting for the delivery of programs and services: from providing unconditional grants to organizations, to using either contracts (purchase-of-service contracting) or contribution agreements that require nonprofit organizations to produce concrete deliverables that can be specified in advance, and are thus in many ways virtually indistinguishable from contracts (Phillips & Levasseur, 2004: 453).

Indeed, the changes that have most often characterized as ‘devolution’ in the literature have been financial in nature. Devolution has involved a shift from long-term supports for operational funding or grants, to short-term, contingent contracting and contribution agreements. These contracts have involved a loss of core funding, and explicit expectations for agencies or communities to make ‘in-kind’ contributions, or obtain joint funding for initiatives or programs by other levels of government (Phillips & Levasseur, 2004; Scott, 2003a; Hall & Reed, 1998). The new contracting processes have also involved increasing accountability demands, and the pre-specification of products and contract deliverables.

The process of devolution has been accompanied by a complementary rationale that emphasizes increasing democratization of policies, programs and services, local relevance, and customized implementations by and for communities. This ‘spin’ on the rationale for
devolution has not simply been hollow rhetoric; it clearly was intended to represent an effort to reverse some of the more egregious negative impacts of modernization – and the loss of community and personal contact that the “business of helping” experienced through being taken over by governments and other secular institutions, as well as the professionalization undertaken in many helping occupations over the past century. Even so, the bottom line for all such discussions has been the overarching rationale of financial efficiency, and reducing costs in the interest of long-term sustainability of programs and services.

From the perspective of the nonprofit and voluntary sector organizations that have been taking on the task of delivering programs and services, this process of devolution has held both promise and peril. Government contracts were initially seen by many organizations as a potential source of stable funding for the longer term, and an answer to perennial challenges of fundraising to be able to address the issues they perceived to be important. The contracts also were perceived as a form of vindication; a recognition that they were better able to identify and meet the needs of individuals who kept falling through the cracks in the government delivered system. Government bureaucracies were described as inefficient and focused on finding ways to limit services and expenditures through eligibility criteria and program regulations – not just by the political critics who were demanding reform, but also by the nonprofit sector advocacy organizations representing the interests of clients. The sector actively promoted a perspective that community agencies had a better on-the-ground sense of client needs, and could deliver better services, and do so more efficiently.

And yet the peril that exists for nonprofit organizations is very real. They risk cooptation of their mandates, increased dependence on government funding to be able to keep a stable work setting for staff and consistent programs for clients, and new
accountability requirements that demand increased professionalization of staff and administration. In short, devolution heightens the tension between nonprofit organizations’ accountability expectations and their autonomy.

**The Impacts of Devolution on Nonprofit Organizations**

The recent literature on the devolution of programs and services has begun to document and highlight a variety of impacts that have evolved, often in unintended ways, from the changes to the process of contracting. These immediate impacts include increased competition among agencies and with the private sector (Evans & Shields, 2005; Eikenberry & Kluver, 2004; Scott, 2003a), moving organizations away from their core goals and historical emphases – mission drift (Evans & Shields, 2005; Eikenberry & Kluver, 2004; Scott, 2003a; Brock, 2000), making it more difficult for community level agencies to advocate on behalf of clients (Evans & Shields, 2005 & 2002; Eikenberry & Kluver, 2004; Scott, 2003a; Brock, 2000) funding and capacity stresses on human resources, such as difficulties in doing long-term planning, and staff turnover (Eikenberry & Kluver, 2004; Scott, 2003a; Alexander, 2000; Evans & Shields, 2002; Hall & Reed, 1998), and longer term impacts on the sustainability of programs and organizations (Evans & Shields, 2005; Eikenberry & Kluver, 2004; Phillips & Levasseur, 2004; Scott, 2003a; Alexander, 2000 & 1999; Brock, 2000). Indeed, the most recent emphasis has been on the challenges that devolution raises for the continued viability of civil society, and the ability of the nonprofit sector to play its important role in contributing to public dialogue about the type of society in which we want to live (Evans & Shields, 2005 & 2002; Eikenberry & Kluver, 2004; Alexander, Nank & Stivers, 1999; Hall & Reed, 1998).
When thinking about the impacts of devolution, it is useful to contrast how that impact is experienced by clients, organizations, and social service provision within communities. For example, most emphases in the literature speak to organization level impacts, such as those noted above – volatility in the contracting environment, mission drift, competition among agencies, and reduced opportunities for advocacy. Less often discussed are the impacts for clients – ‘creaming’ of those clients most likely to succeed, with reduced access to programming for those most in need; the loss of personal advocacy by one agency on behalf of services obtained from another; and the depersonalization people experience as they become defined by their characteristics and perceived program eligibilities rather than having their ‘problem’ viewed from a holistic perspective. The same holds true for system analysis, as the literature rarely discusses broader impacts such as widening gaps in the program and service net, as each agency has increasingly specialized roles that do not overlap with those of other agencies; the shift from long-term to short-term interventions, emphasizing goals of reducing client dependence on support; and an overall bureaucratization of the system, whereby each agency has less flexibility to provide person-centred treatment.

Chapter Five will provide a structured approach to understanding the impacts of devolution on nonprofit organizations and other stakeholders, and the impacts and implications for program evaluation as well. But first, Chapter Four will review key trends in evaluation, and assess their relevance for the context of devolution as a site of evaluation activity.
CHAPTER FOUR: THE ROLE OF EVALUATION IN SOCIETY

In the United States and around the world, our practice of program evaluation is shaped not by the discipline of our profession, but by the corporate and accountability notions of our clients and program managers. We are a service profession, dedicated less to the well-being of stakeholders and communities, more to those with funding and a status quo to preserve. And the future seems to hold more of the same.


Khakee (2003) discusses an emerging gap between evaluation research and practice – between the emerging consensus in evaluation theory, and evaluation as it is practiced ‘in the trenches’. He describes,

... the growing convergence in evaluation research towards stakeholder-oriented, communicative, disaggregated and multi-dimensional methods. On the other hand, public agencies still demand of their policy evaluators quantitative, aggregated, (often uni-dimensional) expert products... the increasing gap between evaluation research and evaluation practice poses some major challenges to politicians, policy makers and public sector managers. (p. 349)

While this gap has been growing for the whole discipline of evaluation, when it comes to evaluation in grassroots and small nonprofit organizations, the gap is a veritable grand canyon. In part, this relates to the capacity of nonprofit organizations to conduct evaluation, but it also reflects a certain myopia within the evaluation literature about the actual contexts in which evaluation work is done. In the nonprofit sector, most evaluation work is not done by ‘evaluators,’ and as such remains outside of the purview of the discipline.

Devolving Evaluation

When government employees were the primary deliverers of programs and services, it was intended that evaluation be done regularly and holistically, examining the program as a whole. As program delivery has devolved to third party deliverers in the nonprofit sector, the program evaluation function initially remained as a centralized activity, overseen or at least
contracted by program management and administration within funding organizations – usually federal and provincial governments. More recently, this process has begun to be replaced by a devolved responsibility for evaluation, whereby individual deliverers are responsible for evaluating or having evaluated the programs and services they provide. In the case of larger nonprofit organizations, this may be done by staff dedicated to such activities, or by external consultants contracted to do the work. For small nonprofit and grassroots organizations, this is often passed along to program coordinators to do at the same time as they are delivering the program or service, or perhaps done with some assistance from an external consultant or a volunteer with specific related experience or expertise. All of these developments have profound implications for how program evaluation work is done, and by whom. Hall et al. (2003) examined how evaluation is conducted in Canadian nonprofit organizations, and who is doing it. They note that only 8% of nonprofit organizations say they hire consultants to help with evaluation work, and another 15% use volunteers. This means that for many nonprofit organizations, the evaluation they are undertaking is being done by staff of the agency, and in many cases, program coordinators and deliverers.

Although the literature on devolution is fragmented and just in the beginning stages of “maturation” (Scott, 2005, p. 155), the evaluation literature examining the context and implications of devolution is extremely limited. There is a growing and increasingly sophisticated literature on how to do evaluation in the nonprofit sector, which has two major emphases: 1) research methods considered appropriate for the nonprofit sector, including a variety of agency-developed manuals and resources (Chalmers, et. al., 2001; United Way of Greater Toronto, 2001), and 2) discussions about the rationale for organizations in the sector to do evaluation (Chalmers, 2003; Behn, 2003; Hoefer, 2000; Fine, et. al., 2000). The latter
often amount to exhortations to “do more” evaluation, and that it is essential to build evaluation and accountability into everyday activities of the organization. Indeed, there is a growing body of literature that focuses on resistance to evaluation, which emphasizes how to convince people of the value of evaluation to organizations, and even how to reduce fear and anxiety (Taut & Brauns, 2003; Donaldson, Gooler & Scriven, 2002). A related literature focuses on evaluation capacity building (Compton, et. al., 2002), although it addresses a wide range of evaluation contexts, and is not restricted to the third sector (McDonald et. al., 2003; Cousins et. al., 2003). Capacity remains a significant issue not just for devolution, but also for evaluation, and is examined in detail in Chapter Five presenting the devolution themes.

The evaluation literature that does address issues relevant to understanding the impacts of devolution and the specific evaluation context of smaller nonprofit, voluntary and grassroots organizations usually addresses specific issues, rather than the context as a whole. For example, writers examine such factors as having multiple contract funders (Koppell, 2005; Markeiwicz, 2005; Mohan, et. al., 2002), stakeholder participation and inclusion (Mathie & Greene, 1997; Thayer & Fine, 2001; Mathison, 2001), stakeholder conflict (Abma, 2000b), multi-site evaluation (Straw & Herrell, 2002), and the challenge of evaluating collaborative and community-based initiatives (Hughes & Traynor, 2000; Page, 2004).

The existing literature dealing with evaluation in nonprofit organizations generally focuses on evaluation as a positive activity; that it helps organizations focus on results, it keeps them aware of outcomes and impacts, and it provides opportunities for stakeholders to contribute to program and policy decisions, among other things. Few studies or papers address the possible conflicts and inequalities that have been a result of devolution of
programs and services. Fewer still address how evaluation can contribute to enhancing the problematic impacts of devolution.

One of the forms of professionalization explicitly encouraged by the marketization of the nonprofit sector is improved ‘accountability’ – although like many other aspects of the process of devolution, what the term actually refers to can be problematic. While the New Public Management insists on accountability as a way of imitating the purported successes of the private ‘market’ sphere, in actuality the defining characteristics of accountability being employed clearly reflect the recognition of the nonprofit sector’s special relationship with government. Aucoin and Heinzman (2000) and Phillips and Levasseur (2004) argue that the state tends to focus on control of abuse and providing assurance that resources are used appropriately, rather than viewing accountability as organizational learning. They state that this is the area of accountability most relevant to innovation and change, but also the least emphasized element of accountability, which is “overwhelmingly focused on control and demonstrating that the rules, particularly those related to financial reporting, have been followed” (Phillips & Levasseur, 2004, 454).

Most of those writing about the accountability emphasis of the New Public Management have this same focus – accountability as contract compliance, avoiding abuse, and justifying appropriate expenditures of scarce resources (Good, 2003; Brock, 2000; Martin, 1995; Ferris, 1993). While the public administration, policy analysis and political science literatures have examined accountability almost exclusively through the eyes of the audit culture rather than evaluation, the evaluation literatures have been notably silent on this context, with the exception that various ‘challenges’ are occasionally identified and discussed in terms of methodological fixes that might address them. The article by Phillips and
Levasseur (2004) is among the first to argue that many of the impacts of devolution experienced by nonprofit organizations are due to the over-riding emphasis on accountability in the New Public Management. Although their emphasis is on accountability broadly defined, they do briefly discuss program evaluation. Another key limitation to this work – in common with most of the existing literature examining accountability in the nonprofit sector – is that it only examines the federal contracting environment. It does not address the full spectrum of devolution impacts experienced by nonprofit organizations, and it does not address the context of small, unincorporated and grassroots organizations, which are also experiencing many of the same impacts as the largest nonprofit organizations.

It is clear that both the context and process of evaluation work is changing – not just for large federally funded programs and agencies, but also for small, unincorporated nonprofit, voluntary and grassroots organizations that operate in one location in a local community. Federal and provincial governments still conduct evaluations of existing and large scale programs, often through a tendered contracting process. Yet increasingly, purchase-of-service orders, contracts and contribution agreements contain not just contract compliance accountability provisions, but expectations for program evaluation focusing on performance measurement and outcome assessment. This work is often done by program delivery staff, sometimes by volunteers or administrative staff in larger organizations, and less often by private consultants or academic evaluators. Increasingly, some combination of these options is employed.

Phillips and Levasseur (2004, pp. 455-460) describe five broad ways that the culture of accountability impacts on nonprofit organizations, and most of these relate to the capacity of nonprofit organizations to cope with increased accountability and evaluation demands.
The first of their themes is how the quest for accountability demands significant time and resources, which are often not funded through projects, but intended to be an in-kind contribution from the administrative budget of the organization. This accountability demand for time and resources encompasses preparing, negotiating, managing and reporting on contracts and contribution agreements. The second theme they describe relates to how they see the contracting environment as risk aversive; requiring deliverables to be specified in advance removes flexibility, innovation and creativity from proposals. In order to get funded, agencies submit more conservative proposals than they would otherwise consider, and focus on features that are most easily monitored. This is exacerbated by a context in which multi-year funding is very difficult to obtain, and concrete deliverables must be attained within a one year project mandate, no matter how realistic that might be.

The third theme described by Phillips and Levasseur consists of time delays introduced by increased bureaucratization of the contracting processes. With new layers of approval for funding, significant delays common, and no possibility for retrospective funding even when project approval is not obtained until several months after start-up, nonprofit organizations end up subsidizing projects. In the current environment in which the majority of projects are expected to have numerous project partners, and often co-funders (usually other levels of government), nonprofit organizations can risk losing both community good will or in-hand funding from other sources by delaying projects. This ends up being a massive juggling act.

Such juggling provides a wide range of stresses for staff in nonprofit organizations, who have no guarantees that funding will come through, which contributes to rapid turnover of staff (and turnover among public servants as well, who can quickly tire of the ‘policing’
role that the new accountabilities demand of them). Even for those staff who remain with the nonprofit organizations, there is a high degree of burnout from working to meet the needs of contracts that do not fund central activities demanded of them, including accountability, evaluation, project development and proposal writing. To complicate matters even further, staff turnover (in both nonprofit organizations and in government) leads to a loss of corporate memory – what projects were done, by whom, and how they were conducted.

Finally, Phillips and Levasseur describe the newest pressures – measuring outcomes, no matter how relevant or easy the process is for the project at hand. Staff are thus addressing such needs without sufficient expertise, off the sides of their desks (as the contract does not officially compensate this work), and usually with insufficient resources combined with unrealistic expectations. In such a context nonprofit organizations often try to keep things as simple as possible; they submit ‘safe’ proposals, and keep the innovation and creativity either underground, or out of the picture entirely.

The impacts of accountability identified by Phillips and Levasseur are relevant to understanding the reciprocal impacts of devolution on evaluation, and evaluation on nonprofit organizations, but they are not enumerative. The range of people who are doing evaluation activity is expanding, as is the nature of the work itself. Increased competition among community agencies also increases the challenge of establishing cooperation among the agencies for conducting research and evaluation, establishing steering committees, and generally implementing inclusive approaches to client support. Having multiple funders brings multiple accountabilities – often with conflicting reporting requirements and reporting overload. From the client and agency perspective, the reporting that gets done with local level evaluations often involves thick descriptions of program activities, and with small
programs this can entail serious risks of lost privacy for clients and staff. The obverse side of this issue is that with local evaluations being conducted, the reports are not necessarily treated with the same criteria as provincial and federal level documents – they may not be made public. This introduces a system level impact of devolved evaluation – a reduction in the transparency that has recently been attained within the evaluation community, whereby official evaluation reports are public documents, and thus cannot be simply shelved if program management or funders do not like the results.

This reduction in transparency is closely linked to another key impact of devolved evaluation – that local agencies can lose the opportunity for dialogue and discussion about their experience delivering programs and services, and how it compares with the experience of other agencies. The community loses the macro perspective of how the program or service links to the broader network of program and service delivery, exacerbating the impacts that the ‘silos’ of individual ministries and government departments have been trying to address over the past two decades of activity aimed at re-inventing government.

The evaluation process itself has impacts for programs and nonprofit organizations. It can rigidify program definitions and eligibility criteria, and make it challenging to provide flexible supports to individuals. This is further exacerbated when funders identify program goals and criteria for measuring outcomes – it intensifies the tendency towards mission drift within agencies.

While staff burnout and turnover obviously have an impact on the process of conducting evaluations, the process of doing evaluation work has also begun to compensate for this situation, as the evaluation work itself, and in some cases the external evaluators have become the new repositories of corporate memory for nonprofit organizations. By
documenting programs and decision-making processes, evaluation work can provide a long-term record for a rapidly changing staff complement. It can also address its official rationale – helping organizations to question what they want programs to do, and determine whether their activities are actually helping them achieve their goals. Evaluation still has that capability to ask the broad question that addresses context and program rationale, but given the challenging context, whether that actually happens is not guaranteed.

**Defining Accountability**

As Phillips and Levasseur noted, accountability is perhaps the most consistently identified point of intersection across both the devolution and evaluation literatures. With the growth of contract-based funding, expectations that contractors should be held accountable for what they are delivering have grown far beyond the expectations that existed in the era of grants-based funding of nonprofit organizations (Phillips & Levasseur, 2004). In the literatures on devolution, accountability has come to represent the vertical hierarchical relationships of the New Public Management; a tool of hegemonic and unilateral control. While this certainly represents a key aspect of what accountability means within the world of nonprofit program delivery, such a portrayal of accountability does not encompass the many subtleties of the accountability dance.

The evaluation literature has held an uneasy truce with the idea of accountability. In a sense, accountability is part of the very rationale for evaluation, which encompasses a diverse range of approaches, purposes and intended uses. Evaluation as a pedagogical activity (Schwandt, 2002), or as a self-learning empowerment approach (Fetterman, 2001) still aims to foster mutual understanding, and help us to identify and discuss the moral and political contexts in which programs occur. This can encompass notions of accountability. Perhaps
more importantly, as Stake (2001) has noted, evaluation is inextricably tied up with the accountability notions of the clients and program managers who contract for evaluation work to be done. So no matter what our intentions regarding the evaluation work we do, those intentions are situated and embedded within specific social and institutional practices (House & Howe, 2000), and as such, subject to the definitions and expectations of others, and not just evaluators.

To be sure, we can ask the question “Is this accountability expectation ‘evaluation’?” Some would say that it is not – and try to make the distinction between evaluation and audit, another uneasy long-term relationship with which evaluators have struggled over time. Yet the overlap and accommodation between the domains of audit and evaluation has persisted because, depending on the context, the line between them shifts, and the line is drawn not just within the two disciplines, but by those who use and contract for this work to be done. Further, while the emphasis of audit tends to remain on ‘outputs’ of activities, this reflects a more short-term pre-occupation, but as audit has grown to encompass long-term implications of those short-term practices, it has begun to focus on outcomes and on the broader societal benefits of programming – traditionally the terrain of evaluation. In many Canadian federal and provincial settings, evaluation activity has been centralized into audit departments, in part as a way of reducing the potential for managerialism possible when the evaluation function is held within the ministry or department overseeing the program to be evaluated.

At a basic level, accountability is about being held to ‘account’ for something, and this implies a relationship with others. Fry (1995) distinguishes between feeling responsible for something compared to being held accountable. We have a ‘freedom to act’ when we are responsible for something, but don’t necessarily have the ‘obligation to answer’ for it, which
we do in a situation of accountability. In this sense we could view responsibility as a wider and more encompassing frame, of which accountability is a subset. Even so, even if we are not being held to account for something we feel responsible for, we can still feel an obligation to communicate about what we are doing to those with whom we interact. Considine (2002) found a high level of consistency across nations in how people felt about responsiveness, obligation and willingness to communicate as forms of accountability among front line officials. It may be useful to think of responsibility as a more subjective end of a continuum, with accountability at the external and outward directed end. Both individuals and organizations can experience this sense of responsibility or accountability.

If accountability implies a relationship to others, we then need to consider who those others might be, and the nature of that relationship. For Mathison and Ross (2002), accountability implies not just a relationship, but interaction, and specifically interaction within a setting that is at once hierarchical and bureaucratic. They discuss accountability as an economic form of interaction flowing from the delegation and dispersal of power within a hierarchical system (their wording implies that this holds both within and between organizations, but this is not their focus); that “those to whom power has been delegated are obligated to answer or render an account of the degree of success in accomplishing the outcomes desired by those in power” (Mathison & Ross, 2002, 2.2). As such, accountability can be seen as a means of controlling both procedures and outcomes in complex settings (Mathison & Ross, 2002, 6.1).

Part of the broader discussion about accountability relates to how the process of delegating responsibility for action can lead to both expectations and a perceived right to an explanation, and also the right to impose sanctions related to achieving those expectations.
And as we move to more complex settings – those involving multiple organizations – the formality of the accountability process increases.

Koppell (2005) examines what he calls pathologies of accountability – he is concerned with the lack of precision in the way that the idea of accountability is used in much of the literature. He develops a typology with five different dimensions or approaches to understanding accountability: transparency, liability, controllability, responsibility, and responsiveness. He talks about how providing such a vocabulary concerning accountability can help reduce conflicts associated with incompatible expectations among those who are demanding accountability, and those who are being held to account. These distinctions are useful for thinking about the rationale for accountability, how it is enacted, and for whom accountability is relevant.

In the evaluation literature, House (1993) relates the drive for accountability with authority structures. He says (1993):

> there are at least three mechanisms for regulating society: the state, civil society, and the market. These are reflected in political (public) authority, cultural (professional) authority, and economic (consumer) authority—power, status, and money. To some degree these forms of authority are convertible into one another. However, they serve different purposes and operate in different ways. Different information is needed to provide accountability and evaluation in each situation. (p. 38)

So for House, accountability in the public sphere is managed through hierarchically organized lines of authority, and in the consumer sphere by the market – people vote with their feet (or their credit cards) based on their perception of the value of goods or services. In civil society, House’s emphasis is on professional associations – a form of member-serving third sector organization that are accountable primarily to their members, rather than to the broader society or to a benefactor.
In his discussion, House treats these spheres as substantially independent. Yet it is clear that with the devolution of human service programs to the nonprofit sector, the civil society model that House uses is becoming less applicable, as professional and public spheres increasingly overlap. As quasi-public organizations, nonprofits are increasingly accountable directly to public ones. In addition, devolution can decrease consumer ‘choice’ in that the processes of the New Public Management usually involve getting rid of overlaps in service, and so there is less choice of which organization to go to at the consumer level.

Warren (2003, p. 50) examines the changes happening within the sector that make accountability an important issue. He outlines a rationale for holding nonprofit organizations accountable that is based on his perception that such organizations are becoming more powerful, which makes their actions potentially dangerous in three ways. The first is that when nonprofit organizations take over responsibilities of the public sphere, they must be held at least to the same accountabilities to the public that we expect of the state – so that we can avoid corruption. According to Warren (2003) the second is that,

… when governments try to capture the virtues of nonprofits by farming out public functions, there is the danger of inequity in the provision of public goods and services. Especially in health and welfare, the general rule is that those locales that are richer in education and wealth also tend to be richer in nonprofits… Nonprofits simply lack the capacities to compensate for inequalities in health and welfare, and so government devolution of public responsibilities to nonprofits can reinforce existing inequalities and cleavages. (p. 50)

And, Warren argues, since nonprofit organizations lack the ability to respond to and compensate for broader societal level inequalities, the state must keep track of things to safeguard the interests of equality. When the state devolves programming to local communities, its process is in many ways a passive one – it relies on the existence of local organizations to deliver programs and services. Communities that lack those organizations may go without, unless the state is able to entice an existing organization to expand to the
community, or support a local grassroots organization to change its mandate or grow to pick up the lacking community capacity. The distribution of nonprofit organizations rarely corresponds nicely with community need, and reliance on such a distribution can reinforce already existing inequities. Sections two and three in Chapter Five address capacity and mandate shift, respectively, and examine some of these impacts of devolution on inequality.

Warren’s third point is that the power that nonprofit organizations gain by becoming major financial players in the delivery of programs can make them influential interest groups regarding public policy. While this is a point of contention within the literature on devolution, his point is clear that with power shifts come responsibilities to account for what is done with that power. This argument is consistent with the previous discussions of Mathison and Ross (2002) and Baulderstone (2005).

Warren’s (2003, p. 51) final point about this shifting context of accountability is that we can examine the nature of the power being transferred, and make decisions about the need for accountability accordingly. If it is easy for individuals to leave an organization – a voluntary, member serving nonprofit, like a club perhaps – then the need for accountability is minimal. If we have little choice in our involvement with the organization – for nonprofit organizations that deliver a service we need in our community perhaps – then it warrants efforts to safeguard the service and our experience with it. For larger organizations, such as professional memberships – even though they may be member-serving organizations – we can look to how significant (large) the resources they can deploy, or how consequential their actions are on peoples’ lives, to see how publicly accountable they should be. So, for example, provincial medical or bar associations serve their members, but their actions and
decisions also have significant import for the health and well being of the population. Thus they need to be publicly accountable.

These models overlap somewhat with one described by Aucoin and Heintzman (2000), which has been widely quoted in the devolution literature (Pollitt, 1999; Boyle, 2002; Phillips & Levasseur, 2004). Aucoin and Heintzman (2000) note:

The purposes that accountability are meant to serve are essentially threefold, although they overlap in several ways. The first is to control for the abuse and misuse of public authority. The second is to provide assurance in respect to the use of public resources and adherence to the law and public service values. The third is to encourage and promote learning in pursuit of continuous improvement in governance and public management. (p. 45)

The third of these purposes is the most clearly applicable to evaluation, although it is the first two that often get the most recognition in the media and the public’s eye. Gregory (1995) focuses on the first purpose, and like Warren, examines the possibilities of corruption, and the influence of both real and potential crises, scandals and tragedies. Similarly, in Canada, David Good (2003) examined the impacts of the Human Resources Development Canada (HRDC) audit scandal from the same perspective, and noted how it changed the accountability landscape for thousands of nonprofit organizations, for which the funding environment became substantially more rigorous, painstaking and demanding.

This introduces another factor into accountability relationships; we do not simply have the relationship between the delegator and the one held accountable, but also the public, or the public’s interest. In fact, as Baulderstone (2005, pp. 10-11) notes, the concept of accountability in nonprofit organizations is quite complex in several related ways. The concept has different meanings for different stakeholders, and depending on the type of organization and its operating environment, the number of stakeholders may be high. While member-serving nonprofit organizations typically have fewer stakeholders, those that provide
public services in a context of devolution often have many intersecting accountabilities, and each of these stakeholders may have different interests, levels of interest, expectations, and standards for accountability. Indeed, standards for nonprofit accountability are often ambiguous, unclear, or non-existent.

How the public perceives accountability further complicates this relationship. Indeed, how those directly involved see the accountability relationships can complicate things tremendously. Mathison and Ross (2002, 2.2) note that “because of the diffuse nature of many hierarchical systems, accountability depends on both surveillance and self-regulation.” Surveillance in this sense represents how those in power hold up their accountability measures as something performed on behalf of the broader society – so they are watchful over the interests of the public, and safeguarding against, among other things, crises, scandals and corruption. But they must then watch over so many programs, services, organizations, and staff members, that this becomes an extremely challenging technical task. So those conducting this surveillance depend also on self-regulation – “the faithful exercise of delegated authority” by those delivering programs and services, whatever their location either inside our outside the public bureaucracy. This is a powerful insight, and will be examined more fully below under several themes in Chapters Five and Six, but here it is relevant to our examination of the public’s ‘right to know.’ Mathison and Ross (2002, 2.3) note that the overall complexity of the accountability relationships serves to obfuscate the interests of various stakeholders, particularly the interests of those in power, who can hide their interests in statements about safeguarding the public good. While Mathison and Ross are primarily discussing this relationship within an internal bureaucratic hierarchical system, Stein (2001, p. 75) describes this process as exacerbated within the context of devolution. Contracting out
programs and services introduces what she calls triangular relationships rather than bilateral, with the public, state and civil society each having interests, and interacting within indirect and unclear lines of authority and accountability. In the process, the transparency of accountability is potentially compromised, and the responsiveness of programs to the needs of the public or the state becomes harder to identify and ascertain.

To summarize our understanding of accountability relevant to the evaluation of programs and services in a context of devolution:

- Accountability involves a relationship between two or more parties, which typically involves ‘economic’ interaction, reflecting the delegation and dispersal of power within a hierarchical system.
- This may be internal to an organization, or involve multiple organizations, and implies an obligation to answer or render an account for actions taken.
- Delegating organizations may expect and have a perceived right to an explanation, and may also have the right to impose sanctions relating to those expectations.
- Accountability can involve transparency – a public obligation to be held to account, reflects authority structures within state, civil society, consumer spheres, can reflect authority structures that span those spheres, and protects us from the dangers of unchecked power implicit in delegated authority.
- Some contexts of delegated power warrant higher accountability concerns, particularly when members do not have the choice of participation, or when the consequences of action are sufficiently consequential.
- Expectations for accountability can include controlling for potential abuse, providing assurance that resources are used appropriately, and promoting learning and continuous improvement.
- Overlapping spheres increase demands for public accountability, while offering more complex authorizing environments, which increase opportunities for obfuscation of goals and interests, and also increase the need for delegating the accountability function itself.

While this summary is not definitive, it introduces many possibilities for discussion as we examine examples of how accountability plays out in a devolved programming context.

**Reflecting on Evaluation and its Rationale**

The starting point for this dissertation is the realization that program evaluation is more than simply applying the latest and most rigorous social science research methods to
the examination of social and educational programs. It is also a political and value-laden activity, and one that is implicated in the official decision-making processes that result in governments spending millions of dollars each year. My emphasis is not on program evaluation methodologies *per se*, although I examine a variety of approaches in view of the larger concerns I am dealing with. I am particularly interested in nonprofit evaluation as a specific political context in which organizations cope with demands for accountability and the changing context of governance under the New Public Management and its aftermath.

The evaluation literature in the discipline over the past fifteen years has begun to move beyond its ongoing methodological preoccupation, and has been addressing the socio-political context of evaluation work in increasing breadth and detail. For example, House and Howe (2000) state that:

> Evaluation always exists within some authority structure, some particular social system. It does not stand alone as simply a logic or a methodology, free of time and space, and it is certainly not free of values and interests. Rather, evaluation practices are firmly embedded in and inextricably tied to particular social and institutional structures and practices. (p. 3)

Evaluation does not just reflect these institutional structures and practices, it reinforces them, and in this way it can have unintended conservative implications. House and Howe advocate for an approach that recognizes the mutually reinforcing relationship between institutions and evaluation, explicitly grounded in democratic principles and deliberation. Their new model, deliberative democratic evaluation, is intended to be reflexive, inclusive, and grounded in dialogue. This model has grown in part as a response to evaluation’s connection to practice – to the design, implementation and evaluation of social and educational programs, and to real world problems requiring real world solutions – it has been proposed as one solution to the issue of values in evaluation practice. As an approach to evaluation that does not represent a ‘method’ so much as an orientation to inclusion and deliberation – and perhaps implies only
the recognition of a diversity of appropriate methods depending on the context – deliberative
democratic evaluation provides “an explicit framework that links evaluation to the larger
socio-political and moral structure” (House & Howe, 2000, p. 3). It asks such questions as:

1. Whose interests are represented in the evaluation?
2. Which stakeholders are represented, and which are missing or excluded?
3. Are there serious power imbalances?

For MacDonald (1976), one of the first to talk about ‘democratic evaluation,’ perhaps
its key concepts are confidentiality, negotiation and accessibility (also noted in Kushner,
2002, p. 3), implicating multiple roles for the evaluator as protector, mediator and proponent
of inclusive approaches. MacDonald’s vision of democratic evaluation encompasses the
rights and obligations inherent in the process – to informing the public, but balanced with the
obligation to protect informants and stakeholders, and to look out for their needs.

The attempt by House and Howe to shift the focus of evaluation ‘theory’ discussions
from their methodological preoccupation is one addressed by Greene (2001c). Greene argues
that for most of its brief existence, evaluation has been a “method-centered enterprise.” Yet
these debates about evaluation methodology have not really been debates about “tools”:

Methods debates are proxies for debates about the nature of the social world and
our ability to know it, and about what’s important to know about the social world,
the role of the inquirer in knowledge construction, and the positioning of science
in society. (p. 1)

When we read the journal articles, conference proceedings and books that make up
much of the official discourse of the professional evaluator, discussions of methods
predominate. Most evaluators recognize the values inherent in the act of conducting
evaluation, yet they have tended to identify issues of values as challenges that can be
addressed by finding methodological solutions. Even so, a growing current of discussion
within the evaluation community has involved attempting to move dialogue beyond methods, in part as recognition that the methods debates do not and can not resolve the issues.

To date, the work on developing a deliberative democratic approach to evaluation practice has been promising. By incorporating a broader appreciation of the political contexts in which evaluation work is conducted, the discipline is developing better tools to understand how evaluation processes can be evaluated and improved. Deliberative Democratic Evaluation appears to be incorporating some of the most salient and practical of the many recent advances in evaluation theory – recognizing the value of input from all stakeholders and a generally inclusive approach to all phases of the evaluation – and seems to hold the promise of becoming an approach that effectively blends many interests and methodologies into a framework capable of cross-cutting the diversity of contexts in which program evaluation is conducted.

Yet as I have noted above, this diversity of contexts is in many ways increasingly anti-democratic. Indeed, I contend that recent impacts of the process of devolution of programs and services have served to make it more challenging to conduct evaluation either deliberatively or democratically.

Evaluation has the potential to be a tool for program improvement and policy development. However the performance measurement focus on accountability to externally determined standards obviates one of the original rationales for devolution: community-level responsiveness. Evaluation in such a devolved setting means that the public transparency efforts that have made evaluation in government increasingly part of the public debate about the worth of programs and the public good might be undermined. It decreases opportunities
for deliberation and debate, particularly among the stakeholders who are most affected by changes to programs and services – deliverers and their clients.

While I am interested in examining factors that might be described as a political economy of evaluation, my overriding interest is in understanding how those broader systemic relationships are constructed on-the-ground. How is it that evaluation practice in nonprofit organizations can aim to improve and rationalize how programs are designed and delivered, and yet can also contribute to the maintenance and even expansion of societal inequalities? How will knowing about the process of reproducing inequalities benefit our ability to introduce systemic change? The approach I employ in the study speaks to this challenge, focusing as a first step on understanding the nature of existing relationships and practice within the context of devolved programs and services.

Stame (2006), in examining the role of evaluation in governance and enhancing democracy, like Khakee (2003) notes the gap between evaluation ‘research’ (theory, in the North American literature), and practice,

… evaluation might aim at enhancing democracy through specific evaluation approaches that reinforced participation, warranted transparency, promoted public welfare… Unfortunately the track record of evaluation studies in this regard is dismal: the debate on utilization has shown that it takes a long time for evaluations to be utilized, and this happens mostly in a cognitive, not instrumental, way; goal displacement seems to be the frequent result of systems of performance measurement; and, in general, public administrators have developed strategies of resistance to being evaluated. (p. 7)

One way that Stame sees evaluation as having responded to these trends is by developing a diverse range of approaches to evaluation that focus on and reflect the diversity and complexity of the operating environments in which it is conducted. Building approaches that allow evaluators to examine vague or contradictory goals (goal free evaluation), and using program ‘theories’ to make the case for evaluation (theory-based evaluation), both
contribute to incorporating the context of evaluation into research design. Indeed, the range of evaluation approaches to choose from, which focus on such contextual issues as defining components of the evaluation process began with Guba and Lincoln’s (1989) fourth generation evaluation, and includes Stake’s (2004) responsive evaluation, Pawson and Tilley’s (1997) realistic evaluation, Patton’s (1994) developmental evaluation, Arnkil’s (2002) ‘emergent’ evaluation, and the growing literature on ‘evaluation use.’ Coping with complexity, and with the unequal power balances within any evaluation context, has become a theoretical concern for the discipline; I believe that nonprofit practice offers opportunities to reflect on the issues for which these approaches have been developed.
CHAPTER FIVE:  
CASE STUDY ANALYSIS: DEVOLUTION THEMES

Whether intentionally or not, involvement in government programs can threaten some of (the) inherent advantages of nonprofit agencies. For example such involvement often creates a tension for nonprofit agencies between their service role and their advocacy role, between their role as deliverers of government-funded services and their role as critics of government and private policies. Such involvement can also put a strain on other important features the organizations, such as their reliance on volunteers, their sense of independence, their frequently informal and non-bureaucratic character, and their direction by private citizens along lines that these citizens think appropriate. Since many of these features are the ones that recommend nonprofit organizations as service providers in the first place, it would be ironic if government programs seriously compromised these features.

Lester M. Salamon, 1989, p. 44.

The range of possibilities for organizing how we think about and understand the effects of devolution on nonprofit organizations is enormous. Most writers in the area emphasize one or more specific themes, and indeed, a problem in the literature to date is that the themes are at many levels of generality, overlap, and in many cases themes from one writer are subsumed within different themes by another. Further, the various themes tend to reflect differing perspectives of impact – on clients, organizations, the nonprofit sector, or society as a whole, and sometimes this happens within one article or theme. A few writers (Eikenberry & Kluver, 2004; Salamon, 2003 & 1997; Alexander, 1999) attempt to develop broader schema for understanding the diversity of impacts. After examining examples from my own practice and the diversity of themes described in the literature, I have chosen five broad themes that appear to be sufficiently distinct for articulation and analysis, although they are by no means lacking in overlaps and caveats, and are not exhaustive of all possible themes. They have been chosen primarily as means for addressing my broader goal of understanding evaluation in a devolved program context, and as such, the goal of understanding and explaining the impacts of devolution is secondary to this end.
This chapter builds on these themes to better understand evaluation in the context of the devolution of program and service delivery. The broad themes are: 1) accountability, 2) capacity, 3) mandate drift, 4) competition, and 5) complexity. The themes are cross-cutting, and some of the examples used are appropriate for understanding multiple themes. This is both expected and intentional.

In this chapter I examine each of the devolution themes in turn. Each thematic section provides three types of information and discussion: 1) examples from the case studies in the form of two or more vignettes – at least one representing a small to medium-sized nonprofit and another representing a grassroots organization, 2) relevant concepts and examples from the literature, particularly focusing on case studies conducted by others, and 3) reflection on the vignettes and the examples and ideas from the literature, representing an initial analysis of the relevance of these concepts for evaluation practice.

**Theme One: Accountability**

Perhaps the most consistently identified issue across both the devolution and evaluation literatures is accountability. Accountability has been a disciplinary constant in discussions and disputes for the evaluation profession since its inception. However, nonprofit organizations have only relatively recently elevated the issue to a primary focus, reflecting the transition from grants to contracts as their primary sources of funding.

**Accountability in the Context of Devolution**

A variety of studies over the past decade have documented how devolution has increased accountability demands on nonprofit organizations. In a survey of 178 nonprofit organizations in the U.S., Fine et al. (2000) examined a variety of evaluation practices. They found that 43% of respondents to the survey indicated their rationale for conducting
evaluation was to meet a funding requirement included in their contract. Larger broker organizations such as the W. K. Kellogg Foundation and the United Way consistently include evaluation and accountability requirements in funding contracts (Scott, 2003a). Poole et al. (2000) note “the drive for increased accountability in human services has put enormous pressure on nonprofit agencies to develop performance measurement systems. A great deal of the pressure comes from government, a chief funding source of services in the not-for-profit sector.” Anderson (2001) used the increased accountability demands placed on nonprofit organizations because of social service devolution as a starting point for examining the use of information developed within those organizations.

Within Canada, researchers associated with the Voluntary Sector Initiative (VSI) have documented an enormous increase in pressures related to accountability and evaluation. The VSI’s 2001 survey of 1,965 voluntary organizations and 322 funders (Hall et al., 2003) found that almost half of voluntary organizations reported an increase in funder expectations over the previous three years. Similarly, nearly half of funders said that they require funded organizations to do evaluations, and another 40% ‘suggest’ that they do so. Respondents also noted a similar level of increase in evaluation requirements by foundations providing funding. The VSI survey included a wide range of nonprofit organization types, and focused exclusively on registered charitable organizations. Those nonprofit organizations that self described as Social Services or Community Benefits organizations were more likely to indicate (42%) that they were doing evaluations because they were required by funders. Interestingly, when asked for the main reason their organizations conducted evaluation, almost three quarters (73%) indicated that it was a decision made by the staff or board, taken primarily for internal reasons. Despite the increase in expectations for evaluation and
accountability measures, the VSI surveys indicate that less than half (47%) of funders provide financial resources for evaluation activities, or allow project funding to be used for evaluation. Almost three quarters of funders provide support in the form of advice. This contrast between high expectations and low levels of support will be examined in the next section on capacity.

What does this demand for accountability look like in the nonprofit organization? Richmond and Shields (2004) argue that current contract financing approaches impose complex and burdensome accountability schemes disguised as evaluation measures. While the study by Hall et al. (2003) showed an increase in expectations about evaluation and accountability, it also showed that the task is also becoming more complex, as organizations try to make the switch from examining outputs to outcomes. The study noted that many respondents from nonprofit organizations confuse outputs and outcomes, and say that they are examining outcomes when they are actually reporting outputs.

A related issue for these organizations is that when they do apply evaluation techniques learned from the literature, it tends to be those that have been designed for business or government, and which do not reflect the nonprofit environment. An example of this comes from contrasting my experience doing evaluation within government, and providing supports to nonprofit organizations.

In my experience working within government as an internal evaluator prior to the devolution of programs to nonprofit organizations, I observed the process of developing a new program. It typically involved having broad political direction given to program management, who involved teams of policy and administrative staff – and occasionally evaluators – in designing programs that bridged the gap between the broad parameters
offered as guidelines from the political domain and the reality of what could be developed and delivered given funding and infrastructure constraints. This process usually involved significant negotiation concerning design, resources, accountability, and responsibility.

After program areas became devolved to third party delivery organizations – in most cases individual nonprofit organizations, but sometimes involving broker organizations such as umbrella groups representing diverse associations of local community level deliverers, a similar kind of discussion took place between program administrators and potential brokers and deliverers of the programs or services in question. Both tried to negotiate with the more central level (for the deliverers, the administrative staff; for the administrators, the political arm giving broad direction) as much flexibility in program design and delivery as possible, trusting their own abilities and awareness of the priorities and client needs so that they would be able to make the best use of funding and resources. Similarly, both of the levels that have any control over funding and resources negotiated as strongly as possible to build control and accountability into their relationship with the respective point of delivery.

When starting an evaluation as an internal evaluator, the first task I took was talking to program management and negotiating a variety of things, including the scope of the evaluation, the range of issues to be addressed, and the timing of program periods and report expectations. One of the most surprising tasks I had to undertake was negotiating the goal of the program. The program staff and management certainly had program goals that they understood, but they were often not written in a format that could be easily shared. Managers liked the flexibility of being able to adapt the program, and they feared that putting the goals in writing would mean that they were being asked to narrow the range of what they could do, and that whatever they said would be used to define their success. If we were not able to
accurately and completely articulate the goals of the program, then we might be trying to hold them accountable for things beyond their control, or worse – inappropriate standards.

When programs were first developed, they usually had some broad goal and objective statements that could be found hidden among program documentation. Initial discussions with program managers and staff to confirm the rationale for the program and the range of activities undertaken under its auspices based on these statements typically began with a laugh and a shake of the head – “those are way out of date,” was a common theme. Program managers had an interest in keeping the program goals and definition obscure, and resisted efforts to clarify the picture of what the program entailed. Part of their rationale for doing so was that the program and context were too complex to be put into a few statements of objectives. Indeed, Abma and Noordegraaf (2003) argue that one of the reasons that performance measurement systems were developed was to simplify complex situations.

The evaluation teams would meet this negotiation process with an approach focused on developing logic models that portrayed the context, goals, objectives, actions, and intended outputs and outcomes of the program. It was a new process for the programs to undertake, and managers thought of it as a useful exercise once they worked through the process of developing one. At this point in time, it was typically done after-the-fact – an aid to framing the evaluation, and not an aid to program definition.

By contrast, the process of program development for those working in nonprofit organizations can be much more rigorous and restricting. The following vignette portrays a typical experience in which I have worked with nonprofit staff to develop programs.
It is late March, and I am in a meeting with Susan, my main contact in the small nonprofit organization. We are in a small meeting room, discussing our plans for the final data collection related to her project, and the unexpected challenges she has faced in obtaining access to teachers in the local school – they are telling her it is now too late in the school year, that everyone is busy, and have asked her to please come back in September. The project ends in May. We are developing our fallback plan.

There is a rapid knock at the door, it bursts wide and a question is flung into the room before the door is half open, “Is the evaluator still here?” Three staff members of the agency are in another meeting room, and they have been hard at work for several days, preparing a proposal related to a “request for proposals” (RFP) tendered by the provincial government for a new program to be delivered by local agencies. They are hoping to get funded and bring the program to their community and agency.

The RFP requests that all proposals contain a detailed evaluation plan, and that the plan include a logic model portraying the program goals, objectives, outcomes, etc. There is an example logic model with the RFP materials, but it is for a very different type of program than they are proposing. They are hoping that I can help them figure out what is required, and review their draft ‘evaluation plan’ to see if it makes sense. It is not part of my official role in working with the agency, but I am almost finished my meeting, and because of the timing of flights to the community, will be in town until the following morning. One of the staff members hints broadly that if they get the contract, they will need the help of an evaluator, so it might be worth my while to help. I tell them that I don’t have any expectation of work from this, but I do agree to help with the proposal. I am interested in seeing what kinds of
expectation the provincial contract has, and how those writing the proposal understand the expectations. And while I think that evaluation is a worthwhile endeavour, I suspect that I respond favourably to their request because of the look of total exhaustion on their faces, and the chaos of the proposal writing room, with empty pizza boxes and a dozen coffee cups and mugs strewn about the area.

The proposal is nearing completion, the evaluation section has been drafted, and we spend a couple of hours working on a logic model. Many of the elements that the RFP suggests should be in the logic model are unknowable at this point – the proposal is for a project that would require a great deal of background work to better understand the nature of the problem in the community, and as a first stage in developing the program, such a logic model and proposal would require that those preparing it already had access to the results of the project’s first phase.

I phone the provincial government contact for the RFP, and ask about the expectations for the logic model and evaluation plan, suggesting that some of the information is not yet knowable. The contact assures me that the review committee understands this, and the draft evaluation plan and logic model are expected to be “a work in progress,” and something that will be adapted over time. I am told that they include it in the RFP because they want to be assured that those writing the proposal and implementing the project “know how to do it.” I relay this information to the proposal development team, who are astonished. For them, the logic model would be part of the contract – it would be what they had agreed to do for the contract. Indeed, for other contracts this has been their experience – once put onto paper, the plan, goals, activities and expectations as first outlined develop a life of their own, and are hard to deviate from. I try to reassure them that a logic model is a tool, and should be
adapted to reflect the program reality, rather than used as an inflexible roadmap. They suggest that the expectation depends on who the project coordinator is for the contract – and that is not necessarily the person who makes the final decision about funding the contract.

**Reflection 5.1.1**

In their national study, Hall et al. (2003, p. 3) note that there is a considerable level of cynicism in the nonprofit sector about the evaluation process – that people have a suspicion that evaluation information is not used, and it is an empty request requiring a filed report that has no long-term significance. I suspect that my efforts to clarify the intentions of government staff only served to confirm these suspicions, although we used the opportunity to discuss why and how evaluation is an important tool for local agencies, whether or not funders expect it. It is interesting that this cynicism co-exists with fears that any evaluations that do not demonstrate positive results will automatically lead to lost funding (Hall, et al., 2003). What is clear from my experience with many nonprofit organizations is that they experience the evaluation and accountability process as a ‘crap shoot’ – sometimes their government contacts are knowledgeable about evaluation and have realistic expectations, sometimes they are rigid and inflexible in their expectations and understanding of what is to be done, and sometimes they start the contract with the former, and finish it with the latter after their contact moves on to a new job.

English and Kaleveld (2003) argue that

... the development of a program logic for the purposes of focusing an evaluation can be a highly politicized process. This is because stakeholders within the authorizing environment (by definition) have a stake in how programs (and their sponsoring organizations) are perceived, and in some cases may build coalitions with other stakeholders to ensure this is a shared view (positive or negative). (p. 40)

Most of the time, the process of developing a logic model portrayal of a program does serve to decontextualize the program, harden the ideas about what the program is trying to
do, and remove the perception that there is flexibility to adapt the program to meet unforeseen or unexpected challenges and client needs. And yet developing such a model often serves an important task in clarifying what people are doing and why, helping them to identify important stakeholders and interests that need to be taken into account. There is a trade-off between the benefits of the clarity obtained by articulating a program in terms of goals and objectives, and putting them into the format of a logic model, and the risk that those benefits will be offset by reducing the real or perceived flexibility of the program to adapt to changes in context, client needs, or implementation challenges. This trade-off is in part a reflection of the process of articulating goals, but the logic model itself is a representation of those goals in a format that becomes crystallized in a public way, and one that can be used in many ways unanticipated by those developing the models.

There are alternative approaches for developing logic models, but they are not yet widespread in use among evaluators, and little understood by those using traditional texts and RFP attachments to learn about the process and purpose of a logic model. For example, Williams (2000) and Fraser (2001) describe alternatives to the diagrammatic form of logic model that involve revisiting goals and processes, and building in an understanding of the program as dynamic, evolving, and in an environment subject to change.

This discussion of how logic models are used in nonprofit evaluation addresses the broader issue of identifying the criteria on which the performance of a program will be based. This is a political as well as practical process, and one subject to negotiation and debate. It is part of the larger movement towards performance measurement that is still sifting down to smaller nonprofit and grassroots organizations. The movement to introduce and apply performance measurement in government and in the nonprofit sector has affected the
contracting process, often in ways that are hard to document. They can affect the perceptions of those in government and nonprofit organizations, even when the perceived expectations do not represent actual policies of the funder.

**Accountability and Performance Measurement**

The evaluation literature has seen a lively debate concerning the use of performance measurement in evaluation. Perrin (1998), in discussing some of the effective uses and misuses of performance measurement in evaluation, describes eight problems and limitations, several of which appear to be particularly applicable to the nonprofit environment. These include:

- Local stakeholders can have different interpretations of the same terms and concepts than do those who are expecting or demanding performance measures,
- Goal displacement – “When indicators become the objective, they result in ‘goal displacement,’ which leads to emphasis on the wrong activities and encourages creaming and other means of making the numbers without improving actual outcomes,” (p. 371)
- Sometimes meaningless or irrelevant measures are used, reflecting the need to have a measure rather than one that is useful,
- Looking at programs out of context and treating them as if they are acting in isolation, which is something that programs never do,
- Sometimes measures are not situationally-responsive to unique contexts, or to changing ones, in which goals are necessarily contingent,
- Measures can lack relevance for decision-making or resource allocation – the kind of situation that enhances cynicism about the process, and
- Performance measurement systems can lead to less, rather than more of a focus on the outcome, innovation or improvement, when rooted in a hierarchical control model.

Lindgren (2001) specifically examines performance measurement in the nonprofit sector. Like Perrin, Lindgren discusses goal displacement, replacing local important measures, using measures that rely on context-dependent interpretations, and which are bound to vary depending on who is interpreting them, and introducing a strong motivation for error in measurement based on program deliverers perceiving that the information being collected is being used to evaluate their performance.
Greene (1999) builds on such themes. In writing about the inequality of performance measurements, Greene makes three key observations. The first concerns the meanings of program quality: “program quality, as a manifestation of human experience, is intrinsically complex and cannot be meaningfully defined by or reduced to simple endpoints” (1999, p. 164), and further, “standards for judgments of program quality are also irreducibly pluralistic and cannot be meaningfully captured in just one rendering” (1999, p. 165). Here Greene is noting that a program’s quality is brief, insubstantial, transient, and multiple – it can be very fleeting, and our experiences of it can be quite ineffable and extremely difficult to explain to someone else. She notes (1999, p. 165):

Judgments of program quality are not meaningfully presented as objective, permanent assessments of a fixed reality, but rather must capture the intersubjective, dynamic, dialogic potential of both the judge and that which is judged.

Greene's second point (1999, pp. 165-199) is about representations of program quality. Here Greene is talking about how we represent our findings and our evaluation work and reports with clients, stakeholders and communities. Standard reports deny the inquirer as a knower, and social inquiry itself as a construction. This is reflected in reports written without reference to the author's voice or position. Among other things, here Greene is referring to how we create the thick descriptions that represent the broad picture of the program reality, which usually gets lost when we write up our evaluation studies.

Greene’s third point (1999, pp. 168-169) deals with conversations about program quality. Here she talks of ‘readerly’ texts – highly structured, controlled, predictable, with precise and clear directions about how to understand the text – versus ‘writerly’ texts, which invite multiple interpretations, and insist that the reader participate in meaning-making, and actively write while reading (referencing Abma, 1997). “Writerly texts seek not closure, but dialogue.”
The logic model (and goal-setting) processes imposed upon nonprofit organizations represent a form of ‘readerly’ text – they try to remove the surprising or the ambiguous, and focus on simplification rather than explication. Abma and Noordegraaf (2003) note this simplification goal of performance measurement – that it is introduced in order to reduce complexity, and in response to uncertainty and ambiguity. They define ambiguity as the absence of, or contradictory interpretations about what needs to be done, can be done, and should be done, as well as when and where they should. They argue that performance measurement may be appropriate when ambiguity is relatively low but it is both difficult and potentially damaging in settings marked by a high degree of ambiguity, which would apply to virtually any small nonprofit or grassroots organization, or most innovative pilot programs.

**Vignette 5.1.2 (Grassroots) – Eleventh-Hour Reframing**

I have been working on the first draft of my final report for a two-year innovative pilot project that is part of a provincial funding initiative. I receive an e-mail from my client Laura, forwarded from the main provincial contact – an evaluation consultant who has been working to coordinate activities among several dozen diverse projects with a community health focus. The e-mail includes an eight-page set of detailed instructions regarding data collection and required report format for the evaluation report.

At the beginning of the project, Laura and I had attended a workshop on evaluation held for representatives of each of the projects funded through this envelope. It had established the provincial contact as a resource to participating organizations – most had no budget for working with an external evaluator, and for other agencies, project coordinators were performing the required evaluation component. I was one of a few consultants who attended the workshop, and I found it encouraging to see that the emphasis was on learning
from the projects, and specifically not focused on ‘accountability.’ Even so, a logic model was required as part of the process; it was developed in draft at the workshop, fleshed out in the weeks that followed, and forwarded to the provincial coordinator as requested.

The eight-page set of requirements sent out near the end of the project represents a shift in emphasis. It speaks of using the initial logic model as the basis for addressing ‘objectives achievement,’ provides a structured reporting format that will allow ‘consistency and comparison across projects,’ and clearly represents an emphasis on accountability and a logic of justification for resources expended.

I am comfortable in adapting the new requests for report format and content to meet the existing evaluation and reporting plan, which has been designed to meet the needs of the organization and the perceived community interests and expectations. I can successfully argue to both the agency board and the funder that the report submitted will adequately and more appropriately meet the project’s intent and provide meaningful information for other communities. In conversation with my client I wonder how those other projects are coping with these last minute change requests, as they may not have the experience to be able to appreciate that the requests represent a shift in emphasis and expectation. Perhaps they expected this accountability emphasis from the start?

**Reflection 5.1.2**

The timing of the request, along with the shift in emphasis from a coordinator who clearly understands evaluation as a learning rather than accountability function, appear to reflect a last minute imposition from provincial government funders. It contradicts information previously shared with projects, and arrives too late in the implementation process to be easily accommodated by new data collection or a changed evaluation emphasis.
It also reflects a de-emphasis on the emergent nature of the innovative projects; that they were intended to adapt and grow and discover ‘what works’ rather than stick to one approach without regard to changes in context, need or formative reflection on the processes being implemented. In effect, the new request appears to represent an after-the-fact performance measurement expectation – one based on locally developed criteria and goals, to be sure – but also imposed as a requirement at odds with values previously articulated by the funder’s representative, which acknowledged the emergent, community-based and complex contexts in which all of the innovative projects have been based.

In this way, the funder’s request serves to “decouple accountability from ownership and responsibility, and therefore revert to accountability as regulation and control, rather than as ‘a vehicle for shared responsibility’ among all stakeholder-citizens.” Greene (1999) further notes that those responsible for identifying performance measures and implementing them are not always the ones who are using them in decision-making, which “severs the crucial link between responsibility and authority” (p. 170). Indeed, it is not clear where the new requests are originating; and perhaps it does not matter, in the sense that the anonymity of the requirement itself is a perfect representation of that severed link.

To summarize this discussion of performance measurement and nonprofit evaluation, many of the issues examined by Greene, Abma, Perrin and Lindgren relate to understanding accountability as control, and how that emphasis may reduce the apparent complexity of evaluation contexts. This accountability emphasis can also have unintended impacts, in part because complexity is an inherent and necessary element of how programs operate. The process of reducing this apparent complexity can also lead to a shifting emphasis – often referred to as goal displacement in the evaluation literature, and as mandate drift in the
devolution literature – which represents a potentially subtle and unacknowledged reframing of program goals that has been associated with the introduction and growth of gaps in service and programming (Phillips & Levasseur, 2004). The third thematic section in this chapter examines the issue of mandate drift in more detail, and the fifth thematic section addresses complexity.

Two other cross-cutting themes that grow out of this discussion are the capacity to meet evaluation and accountability expectations, and the broad question “to whom are programs accountable – funders, community stakeholders, clients, or the broader public?” Ferris (1993) addresses capacity for evaluation by contrasting public accountability with nonprofit autonomy, “The increased role of government contracts and funding of nonprofits has heightened tensions as governments seek accountability and nonprofits seek to preserve autonomy” (p. 363). He argues that governments need to recognize that excessive intrusions into the accountability of nonprofit organizations limit the advantages of having the nonprofit sector involved in program delivery, particularly when the demands are made in the absence of adequate resources to meet the requirements. Capacity is examined in the second thematic section of this chapter.

Many writers have addressed the need for programs to remain accountable to local communities, rather than being refocused solely to the concerns of funders (Alexander, 2000; Phillips, 2000; Clayson et al., 2002; Wallerstein, 1999, Ebrahim, 2002). Certainly in the case of small or medium-sized nonprofit organizations that have their programs rolled into broader umbrella programs of either federal or provincial jurisdiction, this is a particularly crucial trade-off, and one that agencies and community representatives remain aware of in their initial deliberations concerning whether to compete for funding and join with broader
delivery networks. The trade-off for local understanding of priorities for programming can become lost in a relatively short period of time. This is an issue addressed by Campbell (2002) in examining outcomes assessment and what he calls the paradox of nonprofit accountability:

Leaders of nonprofit organizations face a particular bind in responding to the demands for results-based accountability. If they focus only on the project-level outcomes over which they have the most control over which indicators are already available, they risk default on the larger question of accountability to publicly valued goals. On the other hand, if they try to demonstrate the impact of their particular projects on community-wide outcomes, they risk taking credit inappropriately or shouldering the blame for indicators beyond their control. (p. 243)

Campbell also notes that community-wide results take a long time to appear, suggesting that accountability processes need to be based in much longer time frames than is typically allowed or expected by funders. Burnley et al. (2005) also note this concern, which is examined in the fifth thematic section on complexity later in this chapter.

Phillips (2000) examines community responsiveness from the perspective of what kinds of administrative or governance changes would facilitate shifting the emphasis from a hierarchical view of accountability to one focused more specifically on the horizontal governance. One of her concerns is that at present the range of decisions made about evaluation and accountability are ad hoc and vary depending on the jurisdiction, the ministry or department, and even the branches or program streams within departments. She argues that flexibility and responsiveness need to be built into the accountability relationship, which must be applied across departments and as a broad strategy within government, as the ad hoc approach tends to default to hierarchical accountability processes, despite the best efforts or intentions of individuals throughout the system.

Finally, one of the aspects of accountability noted by Koppell (2005) is transparency. This is also an aspect of the trade off between local and hierarchical accountability. As I have
watched the process of devolution of programs and then the evaluation of those programs to nonprofit deliverers, I have noted that it has been accompanied by a potential loss of transparency. Twenty years ago when I worked as an internal evaluator, evaluation reports could easily be shelved and made to ‘disappear’ if they contained findings that were considered to be inconvenient. In the interim, there has been a shift to increased transparency as provincial or federal departments across Canada have come to treat the products of evaluations as public documents, and available to anyone who knows where to look and ask for the reports. Such reports are now routinely available on request or even on-line. Yet the devolution of the evaluation function to agencies and individual programs has introduced a new way for evaluation results to become obscured. The individual reports submitted by agencies and deliverers are often not made available as public documents for privacy concerns – they speak about local issues, name individuals, and use thick descriptions that make it difficult to protect the privacy of clients or program deliverers. As such, they can have their availability restricted, and represent a return to less transparency in evaluation.

Another systemic aspect of the devolution process is the sheer number of community level evaluation reports that are prepared and submitted to agencies and government departments. This has two types of potential impact. One is that with this volume, it becomes extremely difficult to keep track of and reconcile the results of so many piece-meal evaluation efforts. The other is reflected in the lost opportunities to have these evaluation efforts contribute to deliberation about program and service delivery between communities. The accountability streams are narrowly vertical, and unlike when a province-wide evaluation is conducted, the results are not necessarily linked in a cohesive manner. And if
they are linked in some form of meta-analysis, that is typically not shared or done cooperatively with local community participants.

On the issue of transparency, Stein (2001) remains hopeful:

We have arrived at a signal moment in history, when the demands of states and citizens seem to converge, at least in part, although for very different reasons. When the state was the sole deliverer of public goods, it had every incentive as a monopolist to conceal rather than reveal. The new post-industrial state needs transparency among providers, assurances of quality, and evaluations of the effectiveness of the public goods that are delivered. So do citizens. (p. 79)

Perhaps citizens do need such transparency, but devolution appears to offer a potential respite from transparency for those who might be inclined to use it, and this may represent a serious systemic unintended consequence of devolving evaluation.

This section has examined the theme accountability, emphasizing the increased expectations of funders for accountability within the context of devolved service delivery. Examples described how the use of performance measurement and logic modelling processes can serve to obscure the inherent complexity of program contexts, actively reduce the flexibility of program delivery, separate accountability from responsibility, and reduce the flexibility of evaluation and reporting efforts by nonprofit organizations. In this way, program delivery and evaluation can become less responsive to local concerns and implementation contexts, and have unintended consequences at both the community and societal levels. The next section continues this discussion of accountability by looking at the capacity of organizations to deliver and evaluate programs on behalf of government funders.

**Theme Two: Capacity**

This section addresses the capacity of nonprofit organizations to effectively undertake the work of devolved program and service delivery, and to undertake evaluation activities related to this delivery.
Concern about the capacity of the nonprofit sector to take on the work of government devolved to it has been one of the most consistent themes in the literature (Ferris, 1993; Hall & Reed, 1998; Alexander, 1999; Alexander et al., 1999; Wolch, 1999; Fredericksen & London, 2000; McDonald & Marston, 2002; DeVerteuil et al., 2002; Sommerfeld & Reisch, 2003; Mulroy, 2003; Eikenberry & Kluver, 2004; Burnley et al., 2005). More recently, the evaluation literature has also begun to examine the capacity of nonprofit organizations to undertake evaluation activities (Wallerstein, 1999; Botcheva et al., 2002; Hall et al., 2003; Phillips & Levasseur, 2004; Levasseur & Phillips, 2005). Both delivery and evaluation as activities represent stressors on the sector, and contribute to work increasingly being done by nonprofit staff ‘off the sides of their desks.’

Salamon (1993, 1997, 2003) has described the nonprofit sector as growing, resilient, central to the effective operation of contemporary U.S. society, and yet experiencing a state of crisis (1997). Core elements of the crisis that Salamon describes are fiscal and economic – lost or changing income from government, and the marketization of the sector that has involved the introduction of user fees and service charges. The fiscal challenges Salamon discusses encompass the shifts from grants to contracts discussed by Phillips and Levasseur (2004) and others, but also include overall declines in funding to the sector. The Canadian context has not seen comparable dramatic declines in funding, but instead, shifts to the types of funding and identities of the funders (as provinces have taken on program delivery). Further, the marketization emphasis in Canada has been more subtle – particularly in terms of the lower prevalence of introducing direct ‘for-profit’ competitors into the delivery landscape. Even so, both pressures are present in the Canadian nonprofit context.
A key preoccupation of the American and international contexts described in much of the literature is the devolution of responsibility for welfare program delivery. As noted earlier, although some aspects of this delivery has been devolved in Canada – notably employment-related programs and a variety of supplementary programs and services for individuals who are participating in the social welfare system in some way – significant parts of welfare program delivery remain under the jurisdiction of public servants. However, delivery has changed dramatically as individual caseloads have grown for social service workers in government, and specific hands-on tasks involving client support have been devolved to those in nonprofit organizations.

The dynamic that has been most apparent in Canada involves shifts to the contracting process. As open grants have been replaced with targeted contracts, nonprofit organizations have been required to meet a variety of stringent conditions for funding. Some of these, such as the requirement to find and develop agreements with other partner delivery and funding organizations, will be examined in more detail later in this chapter in the discussion of the theme ‘competition.’ Other conditions include the requirement of finding and providing in-kind funding from within the organization or local community, and requirements that any funding provided through the contract not be used to pay for administrative costs within the agency, or for any long-term infrastructure, such as computer equipment or furniture. Such items are often referred to as in-kind, but must be funded from private fundraising efforts, or from grants from other agencies or foundations such as the United Way.

These requirements can be particularly onerous for small grassroots organizations as they entail significant amounts of legwork and community development on the part of those putting together proposals and applying for contracts. Their intent, in part, appears to be a
way of separating ‘serious’ contenders for work as program deliverers from those who may be ‘dabbling,’ as well as ensuring that there is a community-level dialogue and negotiation about resources and priorities. The requirements also ensure that the nonprofit organizations end up contributing substantial time and resources towards the delivery of government programs. As Miller (2005) describes it, if a nonprofit CEO were to transfer to the private sector and have the same limitations placed on them, their experience might be:

You’re now back in the for-profit universe … and you’re the owner of a restaurant. Your paying guest comes to pay the bill, offers a credit card, and prepares to sign the charge slip. But before signing, the guest says, “I’m going to restrict my payment to the chef’s salary. He’s great, and I just want to make sure I’m paying for the one thing that makes the real difference here. I don’t want any of this payment to go for light, or heat, or your accounting department, or other overhead. They’re just not that important. The chef is where you should be spending your money!” (p. 9)

From the perspective of the government funder, the organization already has an administrative apparatus in place, and so there is no need to fund this. From the nonprofit organization’s perspective, this apparatus was capable of handling a smaller agency, and one without the significant burden of a new program, one that possibly requires a new physical space, and takes up time of senior staff, board members and volunteers.

This situation is exacerbated by the requirement of writing detailed proposals and applications, and doing the networking required to ensure a high probability of funding. And when administrative processes within government funders or co-funders are slow or delayed, nonprofit organizations can experience substantial periods in which they are subsidizing government operations, as preparation and implementation must meet external deadlines that are oblivious to the funding processes and timelines of the funding agency (Phillips & Levasseur, 2004, p. 460).
The following vignette illustrates how this dynamic can play out for a grassroots nonprofit involved in developing a proposal to deliver and evaluate an innovative pilot project for a government funder.

**VIGNETTE 5.2.1 (GRASSROOTS) – OFF THE SIDE OF THE DESK**

In my work for a university centre that is coordinating a government-funded demonstration project, I have been asked to help community agencies develop their proposals to participate in the process, and in particular, work with them to figure out how to build evaluation activities into their project proposals. The proposals are the second stage of the process, which involves a solicitation of letters of intent, asking interested agencies to submit a brief letter describing their interest, and answering a couple of questions about the suitability of their community and agency. This process is established in recognition of the significant demands placed on agencies in responding to open requests for proposals, and the idea is to do an initial filter of letters of intent, and then request full proposals from only a dozen communities, of which approximately eight would be funded. In addition, all agencies invited to submit a proposal could apply for reimbursement of a little more than $1,000 towards their costs and investment in putting together the proposal. The project itself requires that participating agencies and communities contribute in-kind resources representing approximately a third of total project costs.

Part of my contribution to developing the proposals involves visiting each prospective site and meeting key individuals who would be responsible for implementing the projects if they are successful. In these site visits I strive to answer questions about which agency and community resources may count as ‘in-kind’ contributions. The funder does not appear to
have a clear idea about what this entails, so the team I am working with tries to be generous and flexible in our definition of what is allowable.

Early in the process, it becomes clear in discussing the proposal development process with the agencies that there is a disparity between the capabilities and resources that different agencies are able to contribute to developing the projects. A few agencies have individuals on staff who specialize in writing proposals – commonly the executive director. A couple of agencies have used the development funding to hire a local consultant to help them prepare their proposals. Most agencies have no staff to undertake such work, so they rely on volunteers – who may be board members, and who may also be working in other agencies or government departments as field workers and program delivery personnel. It is from one of these volunteers that I first hear the phrase “off the side of my desk.” It is explained to me as being so overloaded that there’s no more room to take on additional work ‘onto the desk.’ It signifies the second and third priorities that volunteer and community development work represents to those who are already over-extended in their official capacities and responsibilities. In this context, several individuals describe the evaluation component of the project development to be “off the corner of the side of my desk” – in other words, of such a low priority that it will be attended to rarely, if at all.

Despite the financial support for proposal development and the high probability of success in obtaining funding, several of the invited agencies decline to submit a proposal – the effort represents too much commitment by already taxed staff and volunteers, with a one-in-three chance of failure. One individual describes the situation as, “We just went through this proposal writing process last month, and we put a lot of our hearts and souls into a
project that didn’t get funded. We’re just too burned out and discouraged to do go through that again so soon.”

A total of eleven proposals are received, and eight projects funded. As I contact each agency to begin working on the evaluation components of the projects, I discover that one of the contact people I had met and anticipated working with has moved into another job, and is no longer with the agency or the project team. In another agency, my contact turns out to be a different person than planned, but someone whom I had met in my first site visit. This person is assigned to the project after the contract is awarded, and learns about the agency’s commitment by reading the proposal after-the-fact. This project is one in which the proposal and evaluation plan have been developed by a consultant no longer associated with the agency. The new project coordinator begins our discussion about implementing the project with a clear vision of how both the proposal and the evaluation will have to change in order to be implemented ‘in the real world.’ Negotiation with the agency and coordinator begins.

**Reflection 5.2.1**

By the end of this project – less than one and a half years from start to finish – half of the key contacts for the eight projects have left the agency or the project prior to its completion, and others within the agency take over implementation of both the project and the evaluation activities. Having established evaluation sub-committees in each community, I usually have other volunteers who can take on some responsibilities, but in each case there is a need to bring new people up to speed about the project and the evaluation. For the second phase of the project in eight new communities, a new precedent is established, requesting that each agency identify two contacts – a primary and a backup – and have both attend all meetings in a central community, participate in conference calls, and generally be available
as a backup and alternate resource person. Because we were working with small and grassroots agencies, many of these individuals are volunteers, rather than staff.

The issue of turnover among agencies is a serious one for nonprofit agencies and evaluators to cope with. A contributing factor for the high level of turnover is the contracting process, which can take a great deal of time, and involves a high degree of uncertainty, particularly for grassroots organizations. The volatility is experienced in several ways. When contracts are granted, there is often a very short window in which to hire needed project coordinators. Remuneration is often relatively low, and in smaller communities, the pool of workers available may not be extensive. Some individuals may jump from project to project and agency to agency over an extended period of time, but they are often looking for more long-term, secure, and higher paying work – with benefits if possible. Agencies sometimes have such people contribute in preparing proposals and applying for contracts ‘on speculation,’ with the understanding that they will be hired to do the work if the proposal is successful. However, this means that the projects are also subject to the telegraphed and often clear ‘end-dates’ of contracts. When there is no likelihood of repeat funding for a project, and it ends in a few months’ time, project coordinators are tempted to look to the next project and the next opportunity that comes up, and leave the job prior to the completion of the project – and in many cases before completing the project report and evaluation.

In a Canadian case study of staffing, retention and government funding, Akingbola (2004) examines some of the ways that contract funding and temporary staffing can be detrimental to an agency's services. Akingbola (2004) notes that contract-based funding leads to the hiring of temporary staff, and affects the retention of employees. “Unlike for-profit companies that use alternative staffing for contingent work, the nonprofit (sector) is forced to
use temporary staff for core service delivery in government-funded programs” (p. 463).
Akingbola argues that the consequences of contingent work in the sector include diminished
quality of services, inability to retain acquired program competencies, program instability,
employee turnover, employee distraction, and low employee morale. Even the short-term
projects described in the vignette experienced substantial instability because of the brief
contracting period, as well as capacity stress, regular distraction because of funding issues,
and volatile morale that spanned the highs associated with ‘winning’ the contract, to lows
associated with projects ending or facing the defection of key implementation staff.

Another tension present in the vignette reflects a related facet of the funding game
(Bernstein, 1991) as experienced by grassroots nonprofit organizations. The dance of
negotiating co-funders and community partners and involving them in ‘innovative pilot
projects’ reflects a reluctance to establish long-term funding for new programs (Burnley et
al., 2005). Governments will provide seed money to get things started, to see ‘what works’ or
establish ‘best practices,’ but will not agree to fund projects for extended periods of time.
Indeed, they are often funded with the proviso and assumption that the program, if
successful, will work towards self-sustainability – being able to continue without additional
government funding. Further, these innovative projects must in some way be (or appear)
truly innovative – they cannot obtain repeat funding, and so must continually build upon the
network of connected strategic solutions to ongoing social issues.

For the organizations involved in the pilot project in the vignette, this proposal
development process was complicated by an increasingly common requirement for such
projects – the need to clearly document and evaluate the intervention. While this project was
relatively unusual in that it allowed some of the budget to be applied towards evaluation
activities, and even provided consulting support to this end, it still left agencies with a dilemma. The agencies experienced a trade-off between the need to propose an innovative approach that would maximize their chances of obtaining funding, and yet they wanted to keep the implementation as ‘easy to research’ as possible. Phillips and Levasseur (2004, p. 461) suggest that this trade off between the risk of innovation and the safety of ‘tried and true’ approaches that are measurable and can satisfy funders’ accountability concerns is a common one, and one that often results in reduced innovation in program development. The participants in this project clearly weighed these choices as they wrote their proposals.

Alexander et al. (1999), in a study of the impact of welfare reform and devolution on community-based nonprofit organizations, found that the capacity of smaller agencies to adopt business-oriented approaches to be able to handle government contracts was profoundly limited. In part, this reflects how devolution has been implemented as policies that promote business-oriented approaches, but without the accompanying tools that businesses are routinely able to apply – such as using savings and efficiencies to invest in infrastructure, or paying for overhead out of revenues (Miller, 2005). The financial stress this puts on organizations affects most of what they undertake, including evaluation activities, as the next vignette highlights.

**VIGNETTE 5.2.2 (NONPROFIT) – TRADE OFF**

One of the medium-sized nonprofit organizations involved in the eight-community demonstration project has designed and implemented a very novel approach to community education involving two local schools. I have worked with the coordinator to understand their new model, and to find a way to portray it in the report for the funder. The approach involves the development of a detailed curriculum that I have not seen beyond the initial planning
stages. I am eager to see the final product as implemented and reviewed by the on-site coordinator who has been overseeing the implementation of the project and evaluation. In my third site visit to the community, I meet with the coordinator, Alicia, and express the hope articulated by those in several groups that this innovative curriculum can soon be shared with the other projects, and variations on the approach may be tried with local schools in those communities.

My discussion with Alicia begins on a very positive and collegial note, but quickly takes a most puzzling turn. She tells me that she has great results from the six-month follow-up for the curriculum, which looks extremely promising. But she cannot show me the actual curriculum. I remind her that sharing and documenting the process and the outcomes is a clear part of the rationale for the project, and our contract with her agency. Alicia explains to me that the curriculum itself was developed with that part of the resources considered by the agency as ‘in-kind,’ and so the curriculum belongs to the agency, and not the project or the funder. I express my surprise and concern about the agency’s choice, and my belief that it will affect their ability to obtain the balance of their project funding, as well as future funding through this initiative. I also make an appeal to Alicia’s sense of fairness and collegiality – how important it is to share knowledge about our successes with others. I am told that the decision is not up to her, and that if I want to know more about it, I should talk to Anne, the new and recently hired Executive Director of the agency.

I am able to arrange a meeting with Anne the next day. Prior to the meeting I speak with the project’s director at the university centre, and get direction about the funder’s position on ‘ownership’ of the curriculum. When I meet Anne at a local restaurant, I find her to be pleasant, clearly excited by the potential of the curriculum, but matter-of-fact and firm
– the curriculum was created by her agency’s staff, it belongs to the agency, and they intend
to develop it further so that they can market it for long-term revenue generation. I suggest
that the funder’s view is that they legally own the curriculum, having paid for its
development. Anne’s position is that the funder is unlikely to try to force the agency to
provide the curriculum over a small $30,000 contract – the legal bills would quickly outstrip
the total project funding. Further, she appreciates that she will not be able to obtain the final
contract payment of the 10% ($3,000) held back until project reports are submitted, but it is a
risk that she and the agency are willing to take. And as for future funding, I am reminded that
the funder has a policy of not offering repeat funding, so nothing is really at stake, and she
deems that existing contracts with other departments are secure and not at risk. The meeting
remains cordial; I express my disappointment and hope that the agency will reconsider, but
leave the community convinced that I will not be able to obtain the project information I need
for my final evaluation report.

**Reflection 5.2.2**

Alexander (2000, p. 287) examined adaptive strategies of nonprofit human service
organizations in the face of devolution. In coping with new expectations from funders,
nonprofit organizations have developed a variety of strategies, including: 1) strategic
expansion of services and client bases, 2) networking to stabilize and develop revenue
streams and resources, and 3) increased use of business techniques and technology to
generate outcome measures and an image of effectiveness for funders. In my work with small
and medium sized nonprofit organizations I have seen a growing and virtually unrelenting
focus on trying to stabilize funding, which comes from disparate sources, never covers core
agency costs, and remains short-term, even when the probability of funding being renewed is
very high. Often my initial discussions with nonprofit agency clients – usually the Executive Director, but sometimes board members – revolves around the utility of conducting evaluation as a means of demonstrating to potential funders the capability of the organization, and the effectiveness of the programs and services delivered. The hope expressed to me, particularly with small and grassroots organizations, is that evaluation can help establish or confirm the organization’s legitimacy. While funders are increasingly asking for evaluation as part of the accountability package, being able to demonstrate a willingness and capacity to undertake it is seen as a pre-condition to obtaining even initial contracts with a new funder – a willingness to self-regulate, and competence to be left to do it with minimal input or support.

The underlying common experience of these grassroots and small nonprofit organizations is that they are stretched by trying to take on government contracts, even in an area in which they have been developing and offering their own programs and services in the past. Taking on new contracts leads to a juggling effort to find a balance that will allow them to maintain a consistent and stable staff complement, in part for their clients who develop long-term relationships with staff members, and in part for the staff, who would like some certainty in knowing what their future employment situation might be. The agencies often do not see remaining the same size as a long-term option, and feel pressure to either grow or close shop. But growing slowly or evenly is a challenge.

In the Canadian context, Hall and Reed (1998) address capacity by questioning how much government can download to the nonprofit sector. Their concern is that the sector is diverse but does not have the capacity to handle what is being devolved to it. They note:

The nonprofit sector also has a number of inherent limitations: an inability to generate resources consistently or on a sufficient scale; a tendency to focus on particular
groups of the population, leading to gaps in coverage and the duplication of services; the vesting of influence with those in society having command of the greatest resources; at a historical association with non-professionalized approaches in coping with human social welfare problems. (p. 1)

What are sometimes noted as ‘duplication of services’ and ‘focusing on particular client groups’ are often implicated as problematic by funders, and certainly by the New Public Management, yet the experience of nonprofit organizations can be that these features are part of what makes them most effective as a community-level, ‘bottom rung’ in the social safety net. Alexander et al. (1999, p. 452) suggest that:

Nonprofit organizations play a pivotal role in ongoing efforts to devolve federal government programs and transfer public responsibilities to the local level. In the era of welfare reform, the capacity of social service organizations to serve as the public safety net in a manner implied by devolution proponents has come under question. My case studies and vignettes help to understand part of that mechanism. The transience of programming when funding comes and goes rapidly, organizations experience stress, burnout and turnover, and fluctuate in size and the composition of their staff, means that gaps can rapidly appear. In the past, redundancy in programming among diverse delivery organizations helped to ensure that whatever was happening in one organization would not upset the whole safety net. But as contracts have divided up the available contracting work among community agencies in a more ‘efficient’ and rational manner, the organizations’ capacity struggles and volatility can create extended programming gaps within communities, starting when one agency stops delivering a service, and resuming only after this agency or another has taken on the task once again. To evaluate such a program, with serial delivery by several community agencies, and possibly involving the same delivery staff, is a daunting task if undertaken by someone from outside of the organization or community.

This raises an interesting aspect of how the capacity, turnover and burnout in small and grassroots nonprofit organizations can change my role as an evaluator. When staff of an
agency come and go over relatively short periods of time, the evaluator can represent the only continuity between an early implementation of a program, and subsequent iterations of it. In this sense, my work as an external evaluation consultant can make me a key part of the organization’s corporate memory. For some agencies, I’ve been around longer than most of their staff members, and I’m the only one who can speak to what was done in the past, when proposals were written, and even which other community agencies were involved as partners, competitors or players.

Having worked with nonprofit organizations ranging in size from very small grassroots to medium sized organizations with over a hundred employees working in multiple locations and even communities, I notice that there is a strong correlation between the size and duration that an organization has been in place, and its capacity to take on and effectively cope with contract work in the devolved world of provision of government-funded programs and services. In a case study of some eighteen community-based nonprofit organizations, Fredericksen and London (2000, p. 233) examined possible elements of community-based organization capacity as a way to determine how much capacity exists in the nonprofit sector to take on major challenges for delivery. These include leadership and vision (having a directing board, community participation on and support for the board, vision statements, representativeness of community demographics of staff and board members), management and planning (the existence of formal written policies and procedures for internal operations, evidence of planning, such as a strategic plan, written goals and objectives, a budget), fiscal planning and practice (formal financial statements, organization budget, sources and predictability of funds – self-generated, public contracts, grants, fundraising, revenues), and operational support (predictable levels of staff and skills
among staff, relative balance of staff vs. volunteers, education and training of staff, compensation at a level to attract qualified staff, role of staff in the organization, levels of infrastructure and support, adequacy of physical space, equipment and operational funds). Of the eighteen organizations studied, representing a total of 102 employees, only one exhibited a majority of the elements of organizational capacity that Fredericksen and London were looking for. Certainly in my case studies, such factors are not consistently found or stable within organizations over time, and change based on turnover among senior administrators in the organization, the number and type of contracts in place, and the competitive contracting environment.

Botcheva et al. (2002), in another study examining nonprofit capacity, surveyed twenty-five small community agencies serving children and youth in California concerning their evaluation practices. They found that although aware of the importance of outcomes evaluation, most of the agencies lacked the resources to implement it systematically. They expressed interest in learning about it and held attitudes and beliefs that indicated that they understood that evaluation would be worthwhile to them. Canadian data provided by Hall et al. (2003) support this view of the challenge that evaluation represents for nonprofit organizations, in part by contrasting the rising expectations for evaluation by funders with the lack of resources provided by those funders – the evaluation component is often viewed as part of the administrative apparatus, and is thus expected to be part of the in-kind contribution by the organization.

Another facet of the issue of capacity is the impact that devolution has on nonprofits’ previous sources of revenue. Brooks (2000, p. 211) addresses this in examining what he calls the dark side of government support for nonprofit organizations. He asks the question “does
government funding displace philanthropy, or encourage it?” Through a survey of the literature, Brooks notes a broad pattern of “crowding out,” particularly in social service provision and health. In terms of capacity, the concern is that potential donors to nonprofit organizations are reluctant to give to organizations that are increasingly viewed as a ‘shadow government’ – as part of the broader government program delivery network. It’s like paying extra taxes, just for fun. Overall, he suggests that the claim that government funding stimulates giving appears to lack credibility, and indeed, may do the opposite.

On the other hand, another facet of the impacts of devolution on the capacity of nonprofit organizations may benefit evaluation in some ways. As the delivery of programs and services becomes institutionalized in nonprofit organizations and staff become professionalized, there has been some concern that there are fewer roles for volunteers, and the range of activities in which they can become involved in a nonprofit organization can decline. In the case studies from my practice, I have found that inviting volunteers to participate on evaluation steering committees and workgroups provides a valuable area for contribution to the organization, while also finding scarce resources – freeing up some time of paid staff – and building on the knowledge and skills of long-time volunteers who often have a longer history with an organization than more recently hired professional staff. Volunteers are also more likely to have direct links with other agencies and programs, and as such, bring a broader community perspective to evaluation than do nonprofit staff, who have a perspective on their program that is both more ‘interested’ and narrowly focused.

One final issue concerning nonprofit capacity is how the pinch of resources affects what services are offered by an agency. DeVerteuil et al. (2002) discuss how the devolution of programs and services places greater burden on local resources. Their study examined how
a local community used a variety of strategies to limit, ration, or depress demand for programs in the face of mandated delivery. The strategies were initially articulated through low-level and indirect bureaucratic disentitlements, including quality control and spatial consolidation – reducing access by offering fewer and more remote service centres, rendering them less accessible by increasing travel costs for clients (DeVerteuil et al., 2002, p. 232). Later, disentitlement was achieved by cutting benefits, imposing time limits, and requiring workfare. What I find critical about this example is that it shows how not just the delivery is devolved, but also how the delivery agent at the local level ends up taking on responsibility for determining who gets served and how much. The onus for deciding who can access programs and services shifts down to the nonprofit organizations as they compete with one another for contracts, and ironically, this can also serve to make both the process and the inadequacy invisible to funders. In this way organizations not only self-regulate, but take on the task of regulating clients and shifting from a focus on providing service to those in need, to choosing or identifying those who will be denied service.

This issue spans programs and delivery agencies, and goes beyond the actual time-span of any one program’s delivery period, particularly in a contracting situation. As such, it is not likely to get surfaced in evaluations. The issue is a central part of the next theme to be addressed: how devolution can lead to mandate drift for organizations, and what this can mean for those working to evaluate the overall delivery of programs and services over time.

This section has examined the capacity of the nonprofit sector to take on the work of devolved program delivery, and some of the implications that this capacity stress has for the conduct of evaluation. Such factors as the lack of core or repeat funding, marketization pressures, the need to provide local ‘in-kind’ resources or matching funds from another
source, and the elimination of community-level program and service redundancies combine to elicit stress, burnout and turnover among an increasingly contingent workforce. Such turnover and stress can seriously undermine efforts to provide continuity in conducting evaluation, ensuring that it remains a low priority for agencies and program coordinators. For agency staff, evaluation remains something done “off the sides of their desks.” For someone working as a consultant with and for the organization, the focus on evaluation can be a constant reminder of scarce resources appropriated for a potentially dubious activity.

**Theme Three: Mandate Drift**

This section addresses the stability and continuity of programs delivered by nonprofit organizations in a devolved program and service context. The devolution of programs has been accompanied by structural reorganization of the sector, and a shift in how programming priorities and revenue generation are undertaken within nonprofit organizations. The lure of contract dollars to deliver programs similar to those that organizations are already providing can be difficult to resist. That the new revenues offer opportunities for organizations to grow and stabilize funding seems readily apparent to those in the organizations. That the contract revenues also bring service and delivery expectations that are different from those the organizations have been accustomed to can be somewhat less transparent, and certainly can seem like a detail that just needs to be worked out ‘down the road’. This section examines some of the hidden costs associated with the contracting process, and in particular, how evaluation activities can ameliorate or exacerbate these impacts.

Central to this discussion is the increasing tension in nonprofit and grassroots organizations between their growing dependence on new contract funding, and the lost independence of the organizations in their relationships with communities and clients, and
with respect to their ability to offer unique and targeted programs and services. The vignettes in the section highlight some of the mechanisms for how mandate drift can occur in organizations, and the role of evaluation in possibly magnifying or ameliorating such change.

Although many of the discussions in the literature on devolution focus on pessimistic interpretations of the implications for organizations, Shuman (1998) suggests a potential benefit of devolution – that it raises the opportunity for local organizations to try out new and innovative ways of delivering programs and services. His interest is in the possible role of new information technologies to support decentralized delivery when combined with a variety of forms of centralized and decentralized control. Vignette 5.3.1 portrays a small nonprofit organization that is introducing a new automated reporting system requested by the funder, with hopes that it will meet both local and national information needs.

**VIGNETTE 5.3.1 (NONPROFIT) – THE NEW INFORMATION SYSTEM**

The northern nonprofit organization has ongoing funding from a national department, and also delivers several programs that use funding from several provincial departments. The main focus is on health and health promotion in the community, and the agency has approximately a dozen staff members and twice that number of volunteers who are directly involved with the public provision of programs and services. I have been asked to evaluate the national program envelope, which consists of five broad programming areas, delivered through the central and three regional offices scattered throughout the area in smaller communities. My client is the nonprofit organization, although the evaluation is at the request of the national department, which requires that all agencies conduct such a review every five years.
In my early meetings with agency management, I am directed to use data compiled in an electronic database that has been custom designed for the organization. I am told that this is the third and final version of the software, and has been in place for staff to use for approximately a year. The software has been developed locally to meet the specifications of the national department, but based on an on-the-ground understanding of the local program and service environment. The national department has partly funded the development, with the understanding that if it satisfactorily meets the information needs, it will be licensed for use across the province and possibly across the country – a potential source of revenue for the agency, and a ‘ground up’ approach to information systems development that the department hopes will make it more palatable to administrators in other contracting agencies. The agency executive director presses to have the data from the information system used in the evaluation. She is interested in certifying its usefulness for the national department and prospective licensees in other jurisdictions.

I start examining the database, introduced to it by one of the professional staff who has responsibility for looking after it and for training staff across the agency in its use. Lilly describes the software’s features with obvious pride, taking me through nooks and crannies of how it details participation in agency activities by all clients. She explains that each local office has a computer hooked up to the database through a network, and staff have been taught how to input data about their activities, which is to be done on a weekly basis. As we start examining various data fields, and I make notes about some of the tabulations that I want Lilly to run for me, I start to notice that many of the fields have very low counts, and some are blank. When I point this out to Lilly, and ask her about it, she looks towards the open door of her office, lowers her voice almost to a whisper, and tells me that some staff
have not been enthusiastic about entering data into the system. She explains that they are not yet comfortable with computers, and that they are front line program delivery staff and professional counsellors who may resent having to do what they regard as clerical work.

I quickly realize that there are far less data in the system than the executive director has lead me to anticipate – data entry for several programs started only a matter of weeks earlier, and the oldest data in the system are less than three months old. The consistency of data entry also appears to have been haphazard, at best, and despite Lilly’s enthusiasm, I deem that it is extremely unlikely that the system will provide data useable for the evaluation.

My first efforts to broach the topic of the lack of data in the information system with the executive director are dismissed as a product of the newness of the system, and that once we ask the right questions of it, the data will be available. Hoping that perhaps there is simply an issue of record keeping that can be resolved by devoting some dedicated data entry time by an agency clerical person or my assistant, I start my visits to the regional offices and my initial informal meetings and interviews with delivery staff. I decide to raise the database as an issue for my second stage discussions – after the first group level introductions, and when I meet individuals on a one-on-one basis.

Given Lilly’s portrayal of the possibly computer-phobic response by front line staff, I am surprised to see the apparent facility of staff with using computers to find information relevant to our discussions. They have computers readily accessible, demonstrate a practiced hand in logging into the system and pointing to specific programs and activities I ask about. Without trying to pre-judge why the database has not been used, I start to ask staff general questions about what training they have had, how they use the information system, and what they like and don’t like about the database. While I do not have a mandate to evaluate the
database itself, it appears that I will need to develop a reasonable rationale for not using it, given the expectation by the national department and the executive director that it will have prominence as a central source of evaluation-relevant data for my study.

My discussions with front line staff discover several themes that portray a different picture of the dynamics of introducing the information system. Staff actively resist the database because they feel that it does not accurately reflect the kind of work they do. For example, many of the programs involve delivering information, educational and clinical group sessions, and not just one-on-one counselling. The database does not have a field that allows the staff person entering the data to identify both group and individual interventions, or the kind of group or the number of people attending the group. The database does request individual names for participants – information not necessarily relevant or even known for public education events, and considered by staff to be inappropriate for clinical sessions.

The latter represents another serious concern by front-line staff, who are not convinced of the data security of the database system. Clinical staff refuse to enter confidential information about clients into the system. Counsellors point out that non-clinical staff have access to records, and argue that there are no good reasons provided for putting that information into the databases. The key reason given – that it would allow other (backup) counsellors to access the information if needed, or when clients participated in more than one program – overturns a previously established custom of keeping and sharing physical files in a limited way, and as needed. Counsellors argue that much of the work they do concerns sensitive information, and the confidentiality of that process is vital to their effectiveness. Indeed, some clients had heard about the new information system, and without
knowing anything about it, have refused to talk to counsellors unless they promise not to put any information about them into the system.

As I talk with people across the agency, I hear stories about rivalries and competition among regional offices, mandate and policy disputes between management and staff, and efforts to change the qualifications requirements of staff – particularly those doing clinical work. I also am told that the information system designers provided few opportunities to front line program delivery staff to give input concerning their work. Indeed, when they have had the opportunity to do so, staff seem to have made this the first point of their resistance. Resistance to the database seems to be both overt and covert.

My own evaluation report makes recommendations concerning how to address some of the information system implementation issues, but does not rely on information from it concerning even the most basic program output statistics. The constellation of programs and services portrayed in the organization of the information system does not appear to coincide very accurately with the activities and program definitions provided by staff. In working to reconcile the two, I feel caught between multiple parties with different interests, but also clearly recognize that I lack much of the background necessary to understand all of the rapidly evolving dynamics of the situation.

**Reflection 5.3.1**

Staff resistance to using the database appears nuanced and complex. It reflects a response to control imposed by and perceived to be imposed by the agency’s senior management, the dynamics of conflicts and competition for resources and independence among the regional offices, pre-existing trust issues in relationships among staff, and management efforts to challenge the credentials and required qualifications of staff doing a
variety of clinical work. Efforts to introduce the information system affect the definitions, activities, boundaries and mandates of program areas, which are in dispute. In part, staff resistance appears to represent resistance to participating in self-regulation.

The agency is in a time of transition, and the database represents a focal point for much of that change. The national funding department is using the development of the information system to impose, if not order, a sense of coherence concerning the agency’s programming. They want to understand what is happening, and in order to do this they are clearly favouring and indicating by their requests those activities they view as priorities. The process also serves to polarize positions between staff and management of the agency. In working to reconcile the two, my efforts seem to represent a first step in mediation and negotiation between management and front-line staff – in effect, laying the groundwork for a discussion about what efforts are considered worthwhile, and what is the appropriate work of the agency. Management appears to be using the database as a way to lever change. It is not clear to me whether this change is at the official behest of the national funding department, although the trend does seem to reflect currents happening at other comparable agencies across the province – including those with primarily provincial level funding.

From what I see of the national department’s representative in the two times that we meet, the funder is not averse to having the full range of programming activities incorporated into the database – as long as the key features they are interested in are represented. They seem somewhat unaware of how their requests for specific types of information impact on dynamics within the agency.

Carrilio et al. (2003) observed that those delivering programs and services do not consistently use information systems that would help them collect data required by their
funders, even when they have a great deal of support, technical assistance, and training. Other factors they noticed that influenced the use of such software included organizational leadership, attitudes, accountability expectations of the funder, and the organizational culture and how programs collect, organize, and use data within the organization. Carrilio et al. (2003) focus on the more overt and intended components of organizational funding, rather than on resistance as an underground response.

The vignette highlights some of the potential unintended impacts that information systems, and that even what are intended as benign efforts to capture a picture of what is happening in a program can lead to misunderstanding, the perception of imposed control in the organization, and re-evaluation of basic assumptions concerning what is done, what should be done, and who should be doing it. It can lead to shifts in program and organization mandate. As the evaluator dropped into this situation, even though I tried to keep from taking sides in disputes, my endeavours were used by both parties in the polarized debate to reinforce efforts to direct the changes taking place. By trying to use the information system for the evaluation, I raised what had been an underground area of control and resistance into a public dialogue about the direction of the agency. This discussion was not part of my mandate, but once the genie was out of the bottle, I was obligated to present the information I had compiled in a way that facilitated respectful discussion. Even so, my role in this deliberative process was circumspect and limited – my evaluation work identified issues and provided an opportunity for people to talk about them, but I did not facilitate the discussion.

If I had negotiated taking on that role, I might have focused on the implications of the shifting mandates for the organization. Instead, with my more limited opportunities to input to this process, my report described and contrasted different versions of program goals,
activities, client groups and linkages, as described to me by the funder (expectations), agency management (the new vision), agency front-line staff (what they did), and other stakeholders (clients and community partners describing what they want and need). This presented the organization and staff (who each were to receive a copy of the report), with a framework for identifying how their respective visions of the agency overlapped and differed.

At the heart of these processes of imposed change, resistance and negotiation are pressures to standardize programs and services, or at the very least, standardize the ways in which program and service activities are measured and documented. Such standardization impacts on agencies and delivery staff, but the key impacts are usually experienced by clients of programs and services. The trade-off is often between financial security for the program and agency vs. the client-centeredness of the process as experienced by participants.

Grassroots nonprofit organizations typically become established in relation to some issue of local concern. If they are offering a program or service, it is often in response to a perceived gap among existing programs and services within a community. This local responsiveness and knowledgeability of grassroots and small nonprofit organizations is part of what makes them attractive as potential delivery agents – they have grown out of perceived need, they often have developed based on interagency cooperation and a great deal of community effort by volunteers, and they typically represent innovative approaches to coping with community needs and priorities.

Recent research focuses on how small nonprofit and grassroots organizations that begin to deliver programs and services funded through contracts with government funders experience pressures for standardization. Schmid (2004) examines how nonprofit human service organizations, particularly those serving special needs and at-risk populations, can
lose their unique identities when taking on such contract work. He notes that it affects the
organizations’ roles as gatekeepers to services, and advocates for those most at risk (p. 14).
Similarly, Eikenberry and Kluver (2004), Scott (2003b) and Alexander (1999) identify
comparable impacts on efforts by nonprofit organizations to advocate for their clients. This is
more than simple reluctance to ‘bite the hand that feeds them,’ but reflects what can be a
subtle process of shifting mandate based on trying to “cobble together projects and partners
to survive” (Scott, 2003b, p. 4), and simply working to be a cooperative community
participant in establishing and maintaining credibility for local efforts and proposals.

The following vignette explores some of the subtle ways that organizations can
experience such shifts in the mandates and ‘missions’ of programs and the entire agency. It
describes a variety of the factors that contribute to how the process occurs, and speaks to how
change can happen rapidly and yet in such a subtle way that participants may not even be
aware that it is occurring.

**VIGNETTE 5.3.2 (GRASSROOTS) – STEALTHY CHANGES**

I am meeting the members of a program team. I have been working with the agency
for several years, and have watched as one of their main programs has grown and evolved.
This counselling program has provided long-term supports to community members who are
experiencing multiple problems, and have not found other programs suitable for their needs.
It has been a signature program for the agency within the community – one that has defined
the organization’s place and value as offering unique services that are focused on the
participant’s needs. Indeed, a key element of the program has been the intensive processes
used to help clients identify what they needed, and how to build and plan their next steps.
Three years earlier, the agency obtained provincial government funding related to this program for the first time. Prior to this, the program had been funded through a combination of grants from several sources and local fundraising efforts. The new funding came as part of a broader funding envelope offered through similar agencies across the province, and in order to access this program funding, the organization was required to become affiliated with a network of service providers – a federation or ‘umbrella group’ of agencies offering similar programs and services, and through which this funding is distributed. The funding originates with the province, but the federation of agencies has recently introduced new funding guidelines. These expectations reflect an effort to provide each of the participating agencies with comparisons based on program statistics submitted as part of the reporting process to the funder and federation.

At our meeting to discuss how we could evaluate the counselling program, and what might be some of the appropriate measures for such a program, the cross agency comparisons are raised as an example of what would be problematic for front-line staff. I am told that the average number of clients seen by counsellors in the agency has been lower than the average for other agencies in the comparison reports. While no overt pressure has been put on this or other agencies to increase rates, funding cutbacks loom on the horizon, and the subtext of the comparison process appears to be that those agencies reporting the lowest ‘efficiency’ rates may be at risk of having their program funding cut.

My first reaction is to ask about the origins of these statistics – whether they represent an average per counsellor, or by day of counselling. I am aware that most counsellors in the agency work part-time, and so this would be reflected in any global statistics about clients per counsellor, unless the numbers were standardized in some way. As we work this through,
and look at the comparison tables, it appears that this is likely not an issue, but it is something that the agency’s contact with the federation will pursue to clarify how the statistics are compiled.

I then ask about comparisons of apples and oranges – how does the counselling done at this agency compare with that in other agencies? One answer involves referral to the previous discussion – that even when the counselling is the same, if it lasts longer at one agency, then it might get counted differently than at another. For example, if the statistics of clients per counsellor reflect only the number of clients, and not how often each is seen within a given time period, then this agency is likely to come out less well in comparisons with agencies that focus primarily on short-term counselling. This agency and this program has a history of providing counselling that lasts up to a year in duration. Nowhere else in the community is such counselling available for those who cannot pay a per-session fee. The agency has a short-term counselling program – intended to address immediate needs and rarely providing support for longer than a few visits. But this program’s long-term supports appear to differ from those offered at many of the other agencies in the federation. Although most clients do not actually undertake counselling for the full year of potential support, those few who do increase the average significantly.

At this point, the agency’s new program supervisor suggests that perhaps it would make sense to change the duration of counselling for individual clients to either six or even three months. This would ensure that the agency’s statistics would improve relative to other agencies, but it would also address the issue of ongoing waiting lists to access counselling, and give a clearer rationale for transitioning clients to other supports sooner, rather than leaving clients with the impression that they all had a guarantee of twelve months of
counselling. Shortening this period, even if it could be extended on an as-needed basis, would mean that most clients would be focused on shorter interventions, and not take extra counselling time that was not needed. I note that this represents a change in the agency’s mandate to accommodate the perceived accountability requirements of the funder, and ask whether this shift is one that the agency really wants to make, given its history in trying to provide supports to those in need of long-term counselling. A spirited discussion ensues.

As a final point of clarification, I ask what else the counsellors do besides counselling, because if the statistics are compiled only about hours spent counselling, then other activities undertaken on behalf of the agency should not be used in constructing the statistics. For example, as this is a small agency, most counselling staff also spend some time each week working with crisis support and staffing the crisis line, doing community liaison with other agencies, and providing backup to other agency programs and services.

The ensuing discussion raises the question of how to categorize or ‘count’ participation in counselling groups offered by the agency. Depending on the type of group, these sessions last anywhere from one to three hours, and have between four and a dozen participants. Some are ‘open’ groups, using a drop-in format, and as such, some of the people attending are not part of the agency’s usual counselling clientele. It seems that this facet of the counselling program has been particularly problematic in terms of how statistics are recognized by the umbrella agency, as only ‘new’ clients can be counted, no matter how many may attend each session. I arrange with the agency contact to follow-up this discussion with representatives of the federation – clarifying expectations as well as how statistics are used, and figuring out how the information about group work can be more accurately counted and reconciled with statistics about individual counselling provided by agency staff.
A key pressure experienced by small nonprofit and grassroots organizations is to emulate the larger organizations or risk closing their doors. Taking on contracts for program delivery – even when the programs have been locally developed and established prior to obtaining external funding – introduces a variety of ways that programs can be shaped over time by the funder or its delegated representative. The example presented in Vignette 5.3.2 provides several ways of exploring the mechanisms through which mandate or mission drift occurs. One builds on the incremental changes accompanying the transitory nature of employment in the sector. As people move on to new jobs and different agencies, threads of continuity between community connections can weaken, and become replaced by those established and maintained by paying heed to the funder’s priorities. In the vignette, the agency’s new program supervisor is not aware of the history and rationale for the long duration of counselling offered through the program, and in this way I provide that connection through my historical role and association with the agency. Newer staff only see the current reality of program funding and external requirements, without necessarily being aware of the prior rationales for program features.

Rather than relying on the evaluator to be available to provide this corporate memory for the agency, evaluation can contribute such supports by providing thick, detailed description of program histories and the rationale for programs, and by developing program logics or logic models that provide a clear background for new employees or old who want to understand how and why a program has the form that it does. As noted in section 5.1 on accountability, logic models can serve to reduce a program’s responsiveness and flexibility, and solidify or freeze its features if the logic model is treated as prescriptive rather than
analytically descriptive. Yet logic models can also provide a powerful shorthand overview of the rationale for a program, which can be used to communicate information to funders or community partners, or educate new agency personnel about programs, particularly if the evaluator or whoever develops the model builds in this awareness of the mutability of the logic: that it can change to reflect ongoing program changes and adaptations, but also provide an avenue for building and maintaining continuity for a program and agency.

Another example of pressure to emulate larger organizations can be found in the way administrative convenience can lead to some forms of program monitoring and performance measurement that do not relate to the realities of program implementation in context. This is particularly critical in terms of the way smaller organizations are treated using the same rules established for much larger agencies. For example, the issue of whether staff are part-time rather than full-time impacts agencies on numerous levels: the higher proportionate costs of benefit packages for part-time staff; the increased likelihood of there being part-time employees in a small agency – or rather, part-time in possibly several programs, even when they are full-time employees; and the relatively fixed amount of administrative work associated with any such job, meaning a smaller proportion of the total working hours are devoted strictly to the work of program delivery for those working part-time. These factors affect comparisons of program activities among agencies, and put additional pressure on small organizations to grow and emulate larger ones.

Alexander (1999, p. 62) examines the characteristics of nonprofit organizations that affect their ability to take on devolved programming. Some factors include the size and maturity of the organization – in part reflecting the efficiencies represented by larger and more stable program environments – and the organization’s ability and willingness to expand
into new areas as opportunities arise. Again, this ability to expand reflects the program setting as well – nonprofit and grassroots organizations in small communities can find it very challenging to recruit appropriately skilled front-line staff. This can reflect the shallow skill base of small communities and the need to recruit from larger centres. It can be particularly challenging to recruit for temporary or part-time positions in such situations, although with the impermanence of contract funding, that is the best that many agencies can offer – they need people who are willing and able to arrive and depart on short notice, and who can be quickly up to speed on the requirements of the work. Small and grassroots nonprofit organizations thus can experience a significant disadvantage in competing for contracts, and must look for ways to rationalize program delivery: finding local ‘matching’ funding sources, focusing on services that can best cover marginal costs (and subsequently shifting away from those that are least fundable or are unfunded), and concentrating on those programs and characteristics of programs that are most easily identified, quantified, and represented to funders as outputs or as indicators of successful program delivery. Eikenberry and Kluver (2004) and Alexander (1999) note that such processes effectively increase risk aversion of local agencies, and thus agencies focus on less innovative practices.

Another facet of how nonprofit organizations experience the pressure to report activities in a way that portrays them in a good light relative to other agencies is the organization’s role as a gatekeeper to programs and services for those most in need (Schmid 2004). Alexander (1999) discusses how decisions about who gets service are devolved to local communities in a way that puts the onus on the organization – a subtle process of self-regulation. She gives an example of one director of an

… organization that provides job training indicated that the city funding agency wanted government dollars expended more efficiently. The organization was directed
to serve clients who have a higher probability of success in the program, thereby enabling the organization to demonstrate effective performance. (p. 62)

Such a process of ‘creaming’ the most potentially successful clients is not always as explicit as in Alexander’s example. Consider the example in Vignette 5.3.2, and how the process of shortening the duration of counselling available to clients is likely to reduce services for those most in need. Whether immediately explicit or more subtle, such processes can create a very real conflict for organizations whose official mission is to serve those most in need. For the evaluator, they represent a challenge to see such changes and shifts, as they happen over time and may not be readily apparent within the time frame of a single evaluation project.

At a more structural level, devolution leads to “narrowly prescribed program funding” (Scott, 2003b, p. 4) – compartmentalized programs and services, focused on those attributes that funders perceive to be most important. While local small nonprofit and grassroots organizations have traditionally provided fairly comprehensive services – a kind of one-stop shopping, reflecting a client base with diverse needs and a limited organizational capability for providing specialized programming – government funding envelopes concentrate on specific attributes of service or target client group. In part, this represents administrative convenience, yet it also represents an effort to increase the efficiency of programs by making them concentrate on individual tasks and do them well. The problem with this process as experienced by local communities is twofold. One is simply a matter of capacity – small nonprofit and grassroots agencies provide a one-stop shopping approach because they do not have the size or client base to effectively specialize – which again makes it hard for them to compete with larger agencies. The other problem is experienced both in small and large communities – that this process of compartmentalization of programming belies the real-world complexity of the problems experienced by clients of services.
As unfunded programs and services in both larger and smaller centres are cancelled, the remaining programs shift to acute care and crisis support, trying to cope with the needs of clients who have lost their other supports. For those working on the front lines of agencies large and small, this represents a constant battle to have their actual activities and priorities recognized and legitimated, despite reporting processes that do not reward or acknowledge the activities undertaken on behalf of clients who don’t fit the official program. Alexander (1999, pp. 63-64) describes this challenge as a growth in the number of clients in need of services that have multiple, interrelated problems that are harder to identify and resolve. Perhaps the number of such clients has not changed – maybe they just become more apparent when programs or services no longer provide a breadth of support commensurate with their problems. In larger centres, this can be less transparent because the variety of interrelated and mutually supportive programs and services that are based in large and relatively stable agencies. In small centres such a shift can be dramatic in its impact on clients, when through the vagaries of intermittent funding, agencies, programs and services can disappear in an instant, and leave no alternative available in the community.

Long-term chronic care and prevention work can also be harder to measure, and so it receives less solid funding. In numerous projects I have worked with, front-line staff describe doing prevention and community development work off the sides of their desks, and even when they do obtain funding for such activities, the funding is rarely for a long enough duration to establish a reasonable appreciation of the impacts of this work in the community. In this way the reporting process again shapes which programs are funded and the kinds of programming – emphasizing reactive rather than system-oriented solutions – and this in turn has a proportionately larger impact on small organizations in smaller communities, pushing
more of the community development and prevention work into the voluntary sphere, where the work is also substantially less likely to involve an evaluation component.

Even prior to devolution, the mandates of programs and agencies shifted over time. Yet the rapidity with which programs now evolve, combined with rapid turnover of staff, targeted and standardized programming expectations from funders, and the fragility of links to community partner agencies can make it very difficult to see a coherent picture of how programs and services fit into the community-wide mosaic. As evaluation, too has been devolved, the evaluator is not always in a position to reflect on the larger context of programming, not having access to the diversity of programs, and often bound by confidentiality in dealings with individual agencies. Even so, the evaluator is often in an advantaged position to be able to construct such a community-wide map. One alternative involves re-thinking how the results and findings of local level evaluation is shared among stakeholders. Another involves how community agencies work together and cooperate. The next section on competition and section 6.4 on dialogue and deliberation address these concerns. The final section in this chapter (5.5) deals with some of these issues in examining the complexity of local community programming contexts under devolution.

This section has examined some of the mechanisms for how devolution impacts the mandates of programs and agencies. The work of Schmid (2004), Mulroy and Tamburo (2004), Eikenberry and Kluver (2004), Alexander (1999) and others has brought together a growing body of evidence that the devolution of programs and services increases dependence on government funding, and affects the activities of these organizations – particularly in their roles as gatekeepers to services and advocates for at risk or special needs populations. Some writers have disputed such findings. Rawsthorne (2005), for example, conducted a large scale
survey of Australian nonprofit organizations and concluded that although the organizations had expressed a great deal of concern about the possible impacts of devolution, the reality of it was not as bad as they had feared. My focus on small nonprofit and grassroots organizations in this dissertation, and as described in Vignette 5.3.2 leads me to suggest that Rawsthorne’s data appears to reflect three types of bias that may account for these findings. One is that, like most research on the nonprofit sector, her sample emphasizes and disproportionately samples larger nonprofit organizations that would be better able to weather the storm. Second, her sample thus does not include the many smaller nonprofit organizations that went under – that were not ‘winners’ in the competition for government funding, and that would likely be well able to describe the problematic aspects of mandate shift, lost ability to provide advocacy, and so on. And third, my experience points to how quickly the corporate memory of small agencies changes and is lost with high levels of turnover and constantly expanding and contracting programs and services. The very insubstantiality of the programs and services in question makes it challenging to obtain strong evidence of either positive or negative impacts on organizations and clients. Beyond the issue of sampling, Rawsthorne’s data still clearly demonstrates a climate of organizational stress and ongoing concern and resistance to the contracting process. In effect, she found the negative impacts to be less than anticipated (2005, p. 238), but still unmistakably evident in the sector at large.

Finally, one issue raised in this section has been the growth and potential impact of umbrella organizations that serve as bridges between nonprofit agencies and funders. Brilliant and Young (2004) discuss how the identities of federated community service organizations are also changing. They note that community-based federations of nonprofit
organizations have evolved over the past century; they were originally vehicles for more efficiently collecting and distributing charitable contributions. Over time, their identities have evolved in important ways relevant to our discussion; they have become the gatekeepers for accessing government resources and provide a variety of services to professionalizing nonprofit organizations. Occasionally, they can also serve as the client for evaluation efforts in the sector, and undertake an intermediary ‘quality control’ role on behalf of funders.

The next section will focus on competition and partnership in nonprofit and grassroots organizations, and build on the discussion of the changing role of federations, focusing on the changing relationships among community partners.

**Theme Four: Competition**

This section examines the forms of competition introduced with the new competitive contracting procedures associated with devolution. Such competition encompasses more than simple competition between nonprofit organizations – although that is one focus of this section. The contracting processes can also contribute to competition between communities, and within organizations.

Although this study abounds with ironies, perhaps the most striking is the way new competitive contracting procedures for nonprofit organizations have been built and promoted as a means of developing ‘partnerships’ between state, civil society and local communities. A common mechanism for new competitive contracting processes is a requirement that delivery organizations obtain partners in their work. Such partnerships can take at least two forms. One is a requirement to find co-funders. Most commonly this involves leveraging some form of co-sponsorship of a program or service by another level of government, meaning that the province won’t buy in until the federal government antes up, or vice-versa. It can also
encompass the requirement of agencies or communities finding local co-funders (local
government or another foundation or agency), or agreeing to contribute in-kind resources
towards a project’s implementation costs. This usually means that the agency must do
fundraising or use some money from other contracts or grants towards delivery.

The other form of partnership promoted through the contracting process involves
working with other organizations and agencies to deliver programs and services in the
community. Often this means working with other local organizations to coordinate services
and establish effective referral linkages, but it can also mean linking with broader national or
provincial federations of service providers, through which contracts are administered.

The rationale for introducing competitive contracting procedures appears clear – it
provides the logical foundation for expected cost savings (Van Slyke, 2003, p. 296). In
British Columbia, the 1993 Korbin Commission introduced contract reform to delivery and
contracting, using competitive processes as a way of increasing the consistency of the
contracting process, but also looking for ways to save money in the wake of high federal and
provincial deficits (Burnley et al., 2005, p. 71). And yet as Austin (2003, p. 102) and Van
Slyke (2003, p. 296) have commented, many of the agencies affected by competitive
contracting are in communities in which there is insufficient local competition to make the
process worthwhile. Given the limited resources available by the funder – not all
communities are guaranteed funding – in effect, the process shifts the competition to one
between communities, rather than between agencies within the community.

Vignette 5.4.1 provides a window into the world of the small grassroots nonprofit
organization entering the world of competitive contracting, and some possible implications
for staff, organization, and community.
I have just finished a focus group session with participants in one of the agency’s programs, and the director, Laura, has waved at me to drop by her office before leaving for the day. She laughs as we try to make space at the uncharacteristically cluttered table in her office – it is overflowing with papers and files – and she explains that she is putting together a proposal for funding. In response to my question about the purpose and intended audience for the proposal, Laura explains that this is an unusual proposal, and one that she needs to consider carefully. Indeed, she has been working with the board to consider the implications of submitting the proposal on behalf of the organization. She is interested in bouncing some of her observations off me, and getting an ‘evaluation perspective’ as well.

The provincial government has been working with a federation of agencies to fund a variety of activities across the province – many of which overlap with those undertaken by the agency. At present, the organization has become established in the community based on its advocacy and support for a ‘neglected client group,’ and this constitutes several diverse activities that have been introduced over time – involving individual and group counselling, hands-on advocacy in relation to accessing other programs and services, and public education and information sharing. The activities have not constituted a ‘program’ per se – despite being the activities for which the agency is primarily known in the community, they have been introduced independently over time, and in response to perceived need by agency clients. To date, most program activities have been performed by some combination of volunteers (including board members) and the few staff hired by the agency.

Laura describes the rationale for applying for the funding to me. The provincial federation has been working with the province to expand programming into more
communities in order to ensure that all have access to these resources, but also to establish benchmarks for quality of service where they are provided. The federation has solicited proposals from agencies in the community to deliver the program, and Laura and the board are feeling compelled to submit a proposal. The agency is the only one in the community offering programs similar to those requested in the proposal solicitation. And yet there are two other agencies that will likely submit proposals in an effort to expand their programming into this substantive area if possible. The board and director are concerned about being crowded out of the field if they do not undertake what would be a natural expansion for them, but likely a stretch for either of the other two agencies. The federation will let the contract to some agency, so it might as well be the one most suited to the task. And there is always the possibility that because of their lack of suitability, neither of the other agencies would be successful in obtaining the contract, which would mean that the resources available would go to another community. Indeed, it makes sense to maximize the chance that the community can access this funding, since it is unclear when the opportunity might arise again.

Laura and the board have been discussing the pros and cons of this endeavour. On one hand they see taking on the program as representing a means of stabilizing funding for the agency, which would allow them to provide the level of service they think the community needs. In many ways it seems like a well-timed opportunity for the agency and the community – the organization has been growing, although slowly, and seems like the obvious candidate to take on the program by expanding existing activities.

On the other hand, Laura has been considering a variety of implications for the agency, and about some of them several board members have expressed concerns. Joining the provincial federation would be a major step for the agency, and would entail a level of
growth never previously experienced. The board would likely have to expand, and reach out to other interests and skills in the community – particularly those with financial knowledge relevant to the world of contracting. This could easily change the character of the board, and lessen control over the direction of the agency. Still, it appears to be a trade-off that most board members are ready to take.

Some of the growth that the agency would need to undertake appears to offer more challenges. The program as defined by the province includes some elements that the agency has not yet undertaken. It represents an expansion of some services, and might intrude on the turf of one or two agencies, but the funder insists that only one contract will be let in the community – they can’t split it up across more than one agency. So it’s take all of the parts or none. Another concern is that the agency offers several related types of counselling program that are not included in the Request for Proposals (RFP), and the agency would need to take on this new work. With such an expansion, it may be challenging to maintain those program features that remain unfunded. Finally, and perhaps the main concern expressed so far, is that in order to deliver the program as conceived by the funder, the federation will almost certainly insist that the agency hire professional staff with higher clinical skills and certification than the group of staff and volunteers who have been delivering most of the agency’s services. While this may happen over a period of time, the writing seems to be on the wall that the role of volunteers will eventually change. Although some board members welcome such a change, others will likely resist it quite strongly. Existing staff have expressed a willingness to upgrade their qualifications, but it remains to be seen whether that is a feasible option for some of them.
As we talk, I reflect on some of the experiences of other agencies I have worked with, and the trade-offs entailed by this growth. Although cautious in their approach, the director and board appear to be considering this opportunity with some trepidation, but the overall atmosphere seems to be one of excitement and optimism. Laura asks me whether I can see that they have missed anything. I am tempted to talk about how it is easy to underestimate the potential impacts of such factors as pressures to professionalize and change the nature of programs and services offered, but instead focus more specifically on the evaluation hat I am wearing that day. I ask whether they have done any background needs assessment to determine the level of demand in the community – whether the program is really needed. Laura’s response is that although her information is primarily anecdotal, there appears to be a high level of unmet need in the community, so she is not worried about the issue. I suggest that perhaps a needs assessment as a ‘one-time’ cost might be built into the initial contract – it could be the focus for generating community discussion about how to meet needs, because this programming option might not be the only or even best alternative. She says she will discuss it with the board, and perhaps raise it in her next meetings with other local agencies that offer services to the same client group.

My next comment is about the potential demands that participation in the federation will make on the agency. The reporting requirements may be fairly extensive, and are not likely to decrease over time – they tend to increase, if anything. On this point I am both enthusiastic and cautionary. I have been trying for some time to have agency staff and volunteers improve their record keeping about what they do, for many reasons. I encourage them to be more reflective about what they actually do, whether it is useful for their clients, to solicit feedback whenever possible, and to keep records that can inform their efforts to
solicit funding in the future. And of course, I also am interested in improving the quality of background information relevant to any evaluation work I do with the agency. As I have been employing a participatory approach when possible (and necessary), I have already made some headway on this issue, so I see the formalization of information keeping requirements that the funder will introduce as both possible and as a potential stimulus to improvement. On the other hand, I recognize that the kinds of information requested can be out of touch with the realities of program implementation, and can shift the emphasis of programming. We discuss the need to maintain an awareness of the difference between the measurement and the activity, and Laura expresses confidence that this can and will be done.

Finally, I offer the observation that although the program appears to come with a range of restrictions and limitations, and the funder does not appear to be amenable to alternative ways of defining the program or the local delivery options, joining the federation of agencies represents an opportunity to have a voice about these issues. The agency should be able to participate in the federation in a way that will allow it to advocate for change and recognition of the way they see the problems and solutions in this community. I suspect that there will be an audience for such views, and participation will give them a seat at the table where those discussions are already taking place – but until now, without input from them.

**Reflection 5.4.1**

Rosenberg et al. (2006) suggest that the bureaucratic regulations concerning the competitive contracting process are particularly beneficial to large organizations, rather than small nonprofit or grassroots agencies. Eikenberry and Kluver (2004) note that the processes also favour established organizations with a long-term presence in the field. Mulroy (2003) describes a study of devolution in which smaller community-based agencies perceived that
larger organizations – those without a local neighbourhood focus – seemed to obtain more funding than local agencies by developing a portable program model that they could bring to any community (p. 56). Given these competitive pressures as nonprofit organizations enter into the contracting arena, the last two decades have seen a great deal of turbulence in the sector, with most organizations experiencing pressure to grow, and many others closing their doors after being unsuccessful in bidding for contracts, or over-extending themselves in attempting to grow too quickly or too soon.

McMurtry et al. (1991) discuss strategies employed by organizations trying to cope with declining revenues and increased competition among agencies. They note that the first steps are often small-scale productivity enhancements, cutbacks to services provided to clients, and a variety of forms of agency restructuring (p. 235). They note that it is much more common to find agencies either closing or growing through restructuring than to see mergers or franchising (p. 250).

Franchising, privatization and developing linkages with corporate or private organizations (Eikenberry & Kluver, 2004) are more common approaches in the United States than in Canada, but the most common alternative in either country is becoming part of a federation or umbrella organization that acts as a point of mediation between individual agencies and (usually) government funders. Such federations provide links to other organizations outside of the local community, and according to Eikenberry and Kluver, this aspect of contract competition can weaken traditional relationships, shift the focus to broader issues of ‘the bottom line,’ devalue the work of volunteers, and lead to recruitment of board members with connections to revenue generation rather than within the community (p. 134). The development of linkages with federations thus can be seen as representing a shift in the
form and locus of connection, rather than simply a loss of local connection. One of the challenges of working with such federations is that they can represent both a funnel of funding to local communities, or from the perspective of some agencies and communities, a bottleneck to accessing funding. Federations can promote an approach to program design and delivery that makes it difficult for agencies to obtain funding for anything outside of established and authenticated boundaries of program definition and solution. In terms of evaluation, federations have begun to take on some of the accountability regulation of contracts on behalf of funders, determining which agencies get funded, and for what.

My evaluation emphasis in the vignette is on ensuring that the difficult decisions being made by agency board and director are informed ones. In contributing to the agency’s deliberation, I am interested in helping to recognize the meaningful trade-offs that the opportunity represents. Simply reflecting on whether there is an established need prior to making commitments for a program, checking on the perspectives of and implications for community partners, and establishing the limits of program development and negotiation that can be anticipated for such contracts can reduce agency and staff concerns. On reflection, some issues may warrant significantly higher levels of caution, such as the potential staffing impacts of the push to professionalization, as volunteer efforts are replaced by first staff, and then professionally credentialed workers who are harder to recruit, harder to keep, and usually represent higher costs for the agency. Sometimes these costs are also outside of the control of the agency, when programs funded through federations require front-line delivery staff to be members of one or more unions, and so the agency can be caught between program funding limits and costs that are determined through an unrelated collective bargaining process. A recent example in Nanaimo led to a community agency choosing to
close its doors rather than operate at a deficit, squeezed between unyielding financial pressures (Cordery, 2007, p. A3).

One of my emphases with nonprofit agencies has been to involve volunteers in evaluation work. They may be available and interested when it becomes less possible for them to participate in direct service delivery. Such volunteers often have a wealth of knowledge about programs and the agency, and can be well known in the community. They can carry out much of the legwork of data collection, and provide links to relevant contacts.

Eikenberry and Kluver (2004) emphasize the point that the competitive contracting environment can be detrimental to local community connections – and put civil society at risk. Certainly many authors also note that this process can put pressure on agencies to direct their energies and focus outwards – particularly when it involves establishing linkages to federations or to other partner agencies. Mulroy and Tamburo (2004) suggest that nonprofit managers may need to focus on external relationships – building new collaborations and partnerships. Marwell (2004) describes such an external focus in terms of political awareness – not necessarily political involvement, but certainly being attuned to the political context of programming, and this means both locally and beyond the community. Austin (2003) emphasizes that these efforts will require a new set of skills, encompassing community-building, interagency communications, managing communications, and assessing service outcomes (p. 102).

While Vignette 5.4.1 described a scenario in which the federation and funder were restricting an envelope of program funding to a single agency within a community, a more common situation in British Columbia at present is the requirement that successful applicants for funding have delivery and funding partnerships within the community. Funding
partnerships typically involve other levels of government and/or local stakeholders in providing contributions of resources for projects, and will be examined in more detail in the next section on complexity. Delivery partnerships typically involve local public and nonprofit agencies in coordinating service delivery in some way. As the next vignette illustrates, the competitive contracting environment can complicate such delivery partnerships, and provide contradictory tensions within the web of community relationships. Evaluation efforts can intensify or alleviate such tensions.

**Vignette 5.4.2 (Nonprofit) – Anticipating Opportunities**

A community-based health promotion initiative is winding down after three years of implementation in the community. Early in the project, the coordinator worked with other agencies to develop an inter-agency committee to build linkages among community organizations, including government offices and at least a dozen local organizations that are involved with either co-delivering, referring to or receiving referrals from the program. The group has met on average six to eight times a year, although attendance has declined over the past year. The committee initially undertook several broad-based projects together – a needs assessment, a community protocol for service follow-up, and at least two inter-agency proposals for contracts with government departments. The program will be ending in a few months, and it looks like there is a good chance that the inter-agency committee will either end or go through a transformation as well, since the individual who has been taking the lead in organizing meetings will not be continuing with the project, and likely will no longer remain with the agency.

I met with the committee two times over the duration of the project, and discussed the possibility of forming an evaluation subcommittee that could address a variety of community
issues in addition to the evaluation needs of the project. The message from the committee was that it did not have the broad-based mandate to address much beyond the project at hand, although that might be a ‘good idea’ to look into in the future, if the committee grew into a community fixture. Instead, the committee’s main involvement in evaluation remained at the level of identifying individuals who had information relevant to program evaluation or review, and providing me with an introduction to those individuals if needed.

The key issue I am following up on for the evaluation of the three-year project is referral to the program from community partners. Referral levels for the program have consistently been lower than anticipated, given the initial needs assessment and comments by community partners about the need for the program and size of the target group within the community. As I explore the referral process in my final round of interviews with front-line agency staff of other organizations, I find several process issues that seem to explain some of the problems. One is that some clients are referred, but they do not take the referral – and several organizations do not have any established procedures for following up on referrals. Another is that there has been a significant drop off in referrals as front line staff in other agencies have become reluctant to refer clients to a program that will soon be closing, and for which there are not yet other programs for them to be transitioned to. As I discuss this trend with two of my contacts – both of whom have participated at various times on the inter-agency committee – I come to understand that both of their agencies are in the process of putting together proposals to government to provide a ‘new’ service that would replace the one just ending. One contact explains that there will be a gap in service in the community, and since government does not provide repeat funding for such projects, the other agency cannot offer it again, as they have no alternative sources of funding. But another agency
could step into that void, and try something very similar, but with a few key differences – not
the least of which is that it would be offered by a different agency. The other contact puts it a
slightly different way – that her agency would be a better fit for the program, as most of the
referrals to it came from her agency anyway, so they may as well provide it in-house. Both
suggest that if successful, they would consider hiring the person who is about to be laid off
from the original program to become the coordinator of the new program.

**Reflection 5.4.2**

Although competition among agencies has sometimes been described as a means of
improving service and providing better guarantees that funders are getting value for their
expenditures, it can also be seen as a form of administrative control, as we saw in section 5.1
describe a case study in which the process for obtaining a contract requires an application
form, background information on the organization, and a full proposal (including timelines,
deliverables, activities and expected outcomes), all of which can go through several drafts –
even when there are no other agencies directly competing for this funding (p. 459). And after
the contract is won, there are quarterly reports, cash flow signoffs, final project and
evaluation reports, and possibly an on-site audit. Such a process can encompass completing
forms and paper work consisting of as many as 1,800 individual data fields or components
(p. 459), and can take up to two years for a one-year implementation project – for which only
the one year of direct program activities are compensated by the funder.

In smaller communities where the majority of the proposals submitted to funders are
not related to open bids, this process represents a huge expenditure of scarce resources. The
expectations for the quality of proposals has risen, and so agencies and communities develop
individuals and teams who specialize in proposal development – sometimes as external consultants who work with the agencies. In the face of such a demanding contracting environment, it would seem to make sense that agencies cooperate on project and proposal development, and it does happen. Yet there are several disincentives for agencies to cooperate in this way. One is that even when there is not direct competition between agencies for funding in the community, there is indirect competition – whereby the funder has an allocation set aside for the community or for the region, and once that is committed the funder will not have additional resources for other agencies to apply for. Some communities know that funders like to spread these limited funds around, so they alternate among agencies over time, hoping to ensure that the total of revenues to the community is maintained. Given that some agencies develop stronger ties to funders and are seen as preferred providers, and that in the interest of equity among agencies and communities, funders work to distribute funding among many applicants, the process itself can seem arbitrary or biased. The decision-making processes are not always transparent, and agencies can end up blaming one another or nepotistic ties to other agencies when they are unsuccessful in applying for funding. In such situations the administrative processes can reduce trust among organizations (Mulroy, 2003, p. 56). I have been an observer in decision-making processes in which proposals submitted by various communities are evaluated, and have seen discussions that encompassed evaluating the relative merits, needs and comparative access to funding of different communities.

Another disincentive noted by Mulroy (2003, pp. 55-56) is that the value of individually funded projects does not usually vary to reflect the number of agencies involved in a collaboration. If the resources are the same whether one or five agencies participates,
then the nonprofit organizations are better off going for the full amount, which they would not have to share with others. Contracting in small communities resembles a form of monopsony – a market situation in which only one buyer seeks the product or service of several sellers. No matter the type of service required, suppliers are by definition pitted against one another, and when the amount of resources available to be distributed remains the same or declines, even the closest of allies end up competing for their continued existence.

While many writers note that the competitive contracting environment can make inter-agency collaboration more challenging to accomplish (Eikenberry & Kluver, 2004; Mulroy, 2003; Mulroy and Tamburo, 2004; Austin, 2003), a few have explored the nature of collaboration and the challenges organizations face. Libby and Austin (2002) and Mullet et al. (2002 & 2004) examine coalition building through case studies.

Libby and Austin (2002) examine the process of coalition building among a group of nonprofit agencies in California. The coalition’s stated goal is to develop a comprehensive and seamless system of care for all residents in the area by transforming the way that services are designed and delivered, the way that nonprofit organizations work together, and the role various public and nonprofit agencies. The case study describes the most significant challenges faced by the inter-organizational committee as 1) linking the collective vision to a shared sense of responsibility for services that address client needs, 2) a relationship with funder challenged by communication breakdowns and disagreements, and 3) a constant struggle over the coalition’s identity – shifting from mental health to behavioural health and back again. The study notes that coalition building takes significant investment of time and is helped by having a respected and visionary leader or facilitator, and by being inclusive of as many community organizations as possible.
The inter-agency committee described in Vignette 5.4.2 represents a much more limited form of collaboration, with an agency-specific and pragmatic agenda, but which took advantage of the opportunity to address a few community-wide issues. The inter-agency committee did not undertake to speak with one voice to funders, which was always done through one or another of the agencies, and indeed, the funder may not have learned of the existence of the committee until the final reports were submitted. The committee did not develop a public face – it remained rather *ad hoc* and focused on its narrower agenda with the possible exception of my brief discussions with the committee concerning possible joint evaluation endeavours. Membership in the committee revolved regularly, with both turnover and alternates attending, and consisted primarily of front-line staff whose focus was clearly on implementation issues, rather than broader agency or program policies. A coalition or committee involving senior management of the organizations might have had a different focus, and indeed, some medium-sized or larger communities do establish both such levels of inter-agency committee concurrently. But for many smaller agencies it is a challenge to maintain representation on even one such committee, let alone multiple ones. Building an *ad hoc* inter-agency committee around a specific issue or program implementation effort can be the starting point for a longer term coalition, but such coalitions do not inevitably develop, and can require a great deal of focused attention, resources and leadership among agencies.

The examples by Mullett et al. (2002 & 2004) specifically examine collaboration in the context of the increased vulnerability of agencies facing a competitive contracting environment in British Columbia. They argue that there are few illustrations for community agencies of how to undertake such collaborations, and so they develop a model for creating
local awareness of the competitive structure of funding, and the potential role and value of collaboration in addressing competition. The process

… is intended to create conscientization (Friere, 1996) of everyday practices to examine whether or not they contribute to the community or inadvertently reproduce the ethics of business framework. (2004, p. 163)

They argue that such a new model for collaboration also requires an alternate framework for accountability that embodies the interrelatedness of the agency with the broader community, and that articulates the meaning of collaboration for both the agency and the community. They do not ignore the context of this collaboration, noting that it requires a non-hostile setting in which it can be undertaken.

In addition, there is a particular need for policymakers to strengthen the ability of the voluntary sector to work together to educate funders, to make visible those aspects of their work – beyond service delivery – that are of particular value, and to develop ways to evaluate effectiveness that are not confined to the forms imposed by the methods of the new public administration. (2002, p. 171)

Evaluation that provides thick descriptions of program activities – particularly those that are innovative, non-funded, or part of the broader service provision of the agency but still relevant to program success – can be invaluable in educating funders about the nature of implementation outside the boundaries of what is funded.

My experience as portrayed in the vignette emphasizes that turnover, burnout and other practical issues affect the ability of agencies to develop and maintain collaborative efforts among agencies. Such issues challenge the ability of nonprofit and grassroots organizations to maintain the awareness of the benefits of collaboration that Mullett et al. (2004, p. 163) promote. Further, in the current environment in which nonprofit agencies must deal not just with a single funder, but typically several funders, possibly representing different levels of government, often in ‘partnership’ arrangements and interlocking
programming dependent on co-deliverers and co-funders, keeping track of the many voices relevant to a coalition is a challenge in itself.

Mulroy (2003) addresses the multiple and complex contexts:

… as organizations obtained multiple sources of funding for programs and began implementation, program managers experienced a nonalignment among the funding sources. Each funding source … had its own short time line, regulations, reporting and evaluation requirements, and expected outcomes … This caused constant funding uncertainty that resulted in (1) an inability to plan programs beyond one year, (2) an unevenness in delivering services that required worker continuity and time sensitive interventions, (3) a product for the local community that deviated from what agencies had anticipated in the original grant proposal, and (4) a need for constant grant writing to sustain the implementation of frontline programs. (p. 53)

In effect, such complexity of funders and requirements makes it difficult to identify stakeholders for any individual implementation, which is exacerbated by the need to constantly find ways of borrowing from one program to implement another, and to find common efficiencies wherever possible. Mulroy’s point is that given the growing complexity of implementation contexts, and the seemingly unreasonable timelines for agencies to demonstrate that they are achieving outcomes, the accountability and evaluation efforts also need to adapt by developing evaluative collaborations among providers. Yet as the vignette identifies, when decisions are being made about programs an agency might be offering in the future, participation in evaluations undertaken by other agencies can provide the opportunity to position one agency with respect to another – to anticipate present or future competition among collaborating agencies. Given the potential for agencies working in delivery collaborations to obtain “inside information” about other organizations that may also be their competitors for funding, it seems reasonable to anticipate that it would be worthwhile to develop evaluation protocols and collaboration plans as well.

This section has explored some of the implications of the competitive contracting environment for how nonprofit and grassroots organizations operate. Most writers examine
how competition both demands and challenges participation in partnerships, both locally and more broadly based. Clearly evaluation is implicated in these challenges, as funders use whatever information they have on hand – including the results of evaluation – to help them make decisions about which organizations should obtain or maintain funding, and how they are expected to perform. The next section addresses the complexity of these collaborative contexts for program delivery and evaluation, and Chapter Six introduces more in depth discussion of the nature of collaborative, participatory and inclusive evaluation.

**Theme Five: Complexity**

Perhaps it is a cultural conceit, but in post-industrial society we tend to view small, grassroots, local and rural organizations and communities as rooted in simplicity, and reflecting a clarity of values that harkens back to our distant rural and pre-industrial past. We view our contemporary cities as the ultimate nexus of change, vitality and complexity. In this section I contend that, in terms of the evaluation of programs and services at the very least, small and local communities, and the nonprofit and grassroots organizations that provide services within them, offer environments that are as complex as those in large urban centres at the heart of programming for human services.

In trying to come up with a more useful conceptualization of the idea of community, Shields and Selzer (1997) discuss the concept’s complexities and paradoxes. One of their key points is that even in small, relatively homogeneous communities there can be a great deal of complexity. They feel that Tönnies’ concepts of *gemeinschaft* and *gesellschaft* as a continuum of understanding community are no longer sufficient. Similarly, Barrett (1984) referring to the ‘illusion of simplicity,’ argues that regarding some communities as ‘simple’ serves to hide privilege and power, and reinforces the status quo by contending that in
comparison to central developed areas, small or local communities are lacking something. And yet if we look at the multiplicity of inter-relationships existing in smaller centres such as the grassroots and small nonprofit organizations that are the focus of this study, we see myriad examples of complexity comparable to what we would find in large centres.

In small agencies, individuals wear many hats. A front-line counsellor is likely to also spend a large proportion of his or her time doing such tasks as community development, writing proposals, doing crisis support, answering the phones, and taking out the garbage. In larger organizations it is both possible and more common to find people who have one job, and that is their primary function within the organization – and unless they change jobs, it’s not that likely to change. In my work with larger and smaller organizations, this contrast is apparent when I consider my key contacts in each type of organization. In larger nonprofit organizations they may have one or two people who are given the role of evaluation specialist or liaison. In smaller organizations as I work with staff to evaluate individual programs, I am likely to work directly with each front-line staff person or long-term volunteer over a period of time. Evaluation thus becomes just another hat for staff to put on – it’s part of their everyday task in delivering programs in a small agency. Only the largest organizations have the luxury of specialization of task that can allow some agency workers to develop significant expertise in evaluation. On the other hand, in the small agencies I have worked with, most people get some idea about what evaluation is about, and this can reduce fears, while empowering them to think about their programs in a different way. It also means that my tasks often overlap significantly with education about evaluation and basic research methods, and it means that I spend a lot of time insinuating myself into the details of the daily working lives of front-line staff.
Another way that complexity manifests in the work of small nonprofit and grassroots organizations is that agencies and programs can have multiple funders or sponsors (Boser, 2002; Tang et al., 2002; Fredericks et al. 2002; Guzman & Feria, 2002; Arnikil et al., 2002; Phillips & Levasseur, 2004). Sometimes these are concurrent, and sometimes they are serial. Sometimes they are both. An example of serial multiple funding reflects the way that sponsors provide limited duration funding, with no expectation that it will be extended. When a small agency completes a funded project with one funder and the contract is not extended or renewed, the agency often looks for other funding from another source – another department, or more typically, another level of government. In this way a community project may go through several iterations, each with some differences and adaptations reflecting a requirement that the program reflect some innovation. Few funders are interested in simply taking over a program previously funded by someone else.

As organizations grow – and even when they are first getting into the contract game as a small grassroots organization – they are likely to start offering more than one program or have more than one project going on at a time. These can be unfunded (typically at the start), or each funded separately. Increasingly, agencies are being asked to either co-fund or find co-funders from another level of government (see Phillips & Levasseur, 2004, and Levasseur & Phillips, 2005). The federal and provincial governments do this as a way of leveraging input from other governments, each with the expectation that the other may decide to commit to a longer term of funding if the project proves to be worthwhile. For the agencies involved, it ends up being a balancing act, with one funding source contingent on another, and delays in payments from one meaning that operating expenses may be temporarily borrowed from
another, with disastrous consequences if that funding is cancelled before arriving, or an audit is timed in a way that uncovers the scrambling to keep things afloat.

The evaluation literature has started to examine some of the implications of this increasingly complex web of funding and delivery experienced by community agencies. Bernstein et al. (2002) examine complex contexts in which evaluators need to address the needs of numerous stakeholders, most of whom have divergent priorities. They argue that “It … would be prudent to understand the socio-political environment in which evaluators operate, or what we refer to more broadly as the ‘evaluation authorizing environment.’ Further, we suggest that meeting the needs of diverse stakeholders, given fiscal, political, and logistical challenges, tests the professional skills of evaluators.” (p. 89)

And yet the complexity of contexts brought about by multiple funders and deliverers is not only that there are more stakeholders to be considered – the funders and sponsors have significantly more power than other stakeholders. Fredericks et al. (2002) examine this by noting the increased emphasis on accountability by funders, delegation of service delivery, reliance on nonprofit organizations, and the development of complex social service provider networks (p. 6). While Scott (2003b) describes this as a “house of cards,” Fredericks et al. (2002) provide illuminating detail,

Social service providers … must now rely on a patchwork of public and private funding sources in order to provide services. This has given rise to the development of complex institutional and political networks among social service providers, where formal and informal inter-organizational relationships between funders, program implementers, regulators, evaluators, and the community are often critical to the social service provider’s survival. (p. 9)

Fredericks et al. are particularly interested in the implications of this complex web of program delivery for the field of evaluation. They describe it as paradox – that the complexity of contexts that will make evaluation more challenging is coincident with increasing demands not just for accountability, but for higher quality evaluation efforts:
The results of these trends are that social service providers are confronted with a remarkably challenging and complex environment at the same time that greater demands are being expressed for quality evaluation. This is confounded by the fact that service delivery systems have evolved in such a way that programs are designed across multiple organizations and delivered at multiple sites, such as schools, senior centers, or multiservice facilities, making service delivery and program evaluation even more complex. (pp. 9-10)

Increased demands in a complex multi-funder context enhance such problems as stakeholders with different priorities, competing or conflicting demands for evaluation or reporting, and issues with the development, availability, analysis, ownership and dissemination of data. As a consultant working with such organizations, I have often had to work with a variety of contacts to sort out these priorities, and negotiate expectations from multiple stakeholders. Fredericks et al. (2002) suggest conducting a network analysis in such situations (p. 15), although in the case of working with small organizations with a limited budget, my more common refuge is in putting together some form of evaluation committee that can shift some of the onus for deliberation and negotiation over such issues to the principal stakeholders.

The following vignette describes a situation in which some issues of coping with the demands of multiple funders crept up on the evaluation work mid-way through a project.

**VIGNETTE 5.5.1 (NONPROFIT) – TWO BIRDS WITH ONE …**

I have been trying to make sense of the data from the agency’s internal information system. They have been developing the system for several years, and after a challenging process of negotiation, have reached some compromises with staff about entering program data – particularly individual and group counselling services. The programs are delivered by staff and volunteers out of four regional community offices, and represent a combination of funding from two federal and three provincial government departments. The database has been developed with financial support and at the urging of one of the federal departments,
and the agency’s board and management have been looking forward to having it completed in a way that could simplify report preparation for all programs, as the administrative demands on the agency have grown significantly over the previous five years, and now almost a quarter of the agency’s employee time is devoted to record keeping – including compiling information, entering data, and preparing reports.

The latest round of reports to which I am contributing are attempting to contrast the program and agency clientele and the local population characteristics with the broader community and region on a variety of key health indicators. As we attempt to build the comparison using this database, a community survey that is part of my evaluation effort, provincial government health statistics, and data from Statistics Canada, we come to realize that the census districts do not match up with regional health council boundaries. Further, the survey and database information cuts across both sets of boundaries, as it reflects four geographic communities – but divided two by two provincially, and three and one federally. Neither data set represents a true comparison of the broader population of the regions, and the data are aggregated at a level that does not allow further breakdown without getting expensive customized data tabulations – something out of reach of the agency, and not funded by either level of government, despite the expectation that reports reflect such data comparisons. We begin the process of reconciling our data and negotiating for the next reporting period (five years down the road for the federal programs) to obtain different data collection parameters, reflecting more meaningful boundaries that could meet both federal and provincial information needs. We do this knowing that by the time the next data are collected and compiled, there is a good chance that the agency’s boundaries, constellation of programs, and client base will have changed.
As we start to sort out the various types of information we do have to work with, it becomes clear that within this data the categories and data aggregations requested by federal and provincial departments are not the same. When we have primary data we can aggregate them appropriately for each funder, but we cannot do one report that would satisfy all sponsors. The categories of such fields as educational attainment, age ranges, and First Nations status differ in minor ways, but enough to entail significant added work to sort out the differences and present the data appropriately. When we get to the actual program statistics for the reports, we find that the database is not clear – some counselling falls into several program categories, and we may be double counting some data in several programs, or worse, from the agency’s perspective, not counting some clients that have been served.

**Reflection 5.5.1**

When a contract extends over a long period of time, program administrators and evaluators can work together to find and build efficiencies in efforts to monitor and evaluate programs. When programs are part of a shifting maze of multiple contracts, with turnover among key players, changes to funding regulations, the rise and demise of various co-funders, co-deliverers and community stakeholders, and changes to the linkages of programs and services within the community, there is seldom sufficient long-term continuity to conduct two evaluation studies with anything resembling a comparable methodology, data sources, or for that matter, basic intent.

In the vignette, the federal funder provided the incentive for developing a database that the agency hoped would also serve the needs of provincial program sponsors. While accommodation on such matters occurs regularly, it takes discussion and effort, and is affected by turnover of agency staff and contacts within funding organizations. More
recently, agency federations have also stepped into this maze, developing their own information systems to be used by delivery agencies, and even developing on-line reporting systems for program statistics that go directly to funders. Some agencies I work with must use several such pre-packaged information systems concurrently – each provided by a different program sponsor, department or level of government. The administrative demands can seem overwhelming for those delivering the programs, and for me, trying to make sense of the data so that I can use it for evaluation work is a consistent challenge, as I am usually coping with issues of data quality and access to data – for which I must teach the data administrator what I need and the format of relevant reports or tabulations in the majority of cases in which I cannot obtain access to the unaggregated data.

Anderson (2001) notes that while organizations are collecting more information than ever before, it is usually underused within agencies for decision-making. In part this reflects the gap between the kinds of information demanded by funders, which may not coincide with what program deliverers think is important or relevant to their decision-making. In such situations it is difficult to achieve buy in and commitment to the process by managers. As Boser points out (2002, p. 23), funders typically set the parameters of data collection and evaluation – what is required (defining the appropriate outcomes), when it is required (even when that does not fit with the natural rhythms or timelines of a program’s life cycle), and what resources are available to accomplish the tasks. They represent administrative convenience and priorities, and are often perceived to be arbitrary. Chaskin (2003) further argues that these expectations lack clarity, lack awareness of opportunities or the need for collaboration, and as noted by many others, typically lack adequate resources (pp. 77-78).
One aspect of the multiple funder and multiple stakeholder dance that gets little discussion in the literature is how such stakeholders are not always either clearly identified or unambiguously representing only one stakeholder group. I have found that, particularly in small communities, many stakeholders wear several hats at one time. The next vignette describes an example of how an evaluator can experience this during data collection.

**VIGNETTE 5.5.2 (GRASSROOTS) – SWITCHING HATS**

I have been attending an educational workshop conducted as part of a program that I have been hired to evaluate. I am an observing participant, as I want to understand what the workshops are like for those who are attending – most of whom represent front-line staff of other community agencies. Several young people who have been participants in the program have been part of the presentation to an audience of approximately one hundred people. As they break up into small discussion groups, I join one group of ten people and for a half hour we talk about some of the implications of the presentation, and try to answer some of the questions assigned to the group for reporting back to the main plenary.

One of the young people in each breakout group has participated in the program, and the young woman in my group also has been joined for our discussion by her mother. As we proceed I am very impressed with how articulate the young woman is in talking about the challenges of her situation, and the benefits and deficiencies of the range of programs and services available to her. The woman’s mother also appears to be a strong advocate for her daughter, and knowledgeable about the topic, so at the end of the session I approach them both, explaining that I am working on the evaluation of the program, and that I am interested in following up and interviewing them about their experiences. They readily agree, and as the young woman turns to go back to the front of the room to continue the presentation, her
mother leans over and says to me, “I’m so glad we met. I’ve been meaning to have a word with you – I’m on the board of directors, and you work for me!”

**Reflection 5.5.2**

It is not uncommon for an evaluation to involve interviews of diverse key informants of a program, particularly in small communities and working with grassroots organizations, where the individuals who have developed a program may exemplify the ‘stake’ that we think of when we discuss stakeholders. Such individuals have a strong vested interest in the progress of a program and any evaluation done of it, and they have often had a long history of acting as advocates for both the program and the issue it is addressing. In the vignette, the mother of the young person who has participated in the program is clearly strongly invested in the program’s success. It is perfectly appropriate to interview such individuals for an evaluation, but their orientation and allegiances are also relevant to understanding and contextualizing their comments. In my work in recent years, because of such experiences as that described in the vignette, I have tended to use protocols to ensure that I know as much as possible about the possibly numerous connections that individuals bring to the situation. I assume in advance that they may have many possible hats, and while I cannot know all of them, I can work to identify which hats are relevant to the program and the study.

In the same way that the evaluator cannot know all of the many hats that stakeholders wear, they cannot know their priorities and expectations, which can be diverse even when the role they are playing is well defined. Fredericks et al. (2002) discuss the unpredictability of stakeholder reactions to an evaluation project, and suggest a similar approach to developing a protocol that helps identify stakeholders and inform them of the evaluation and the possibility of participating in it (pp. 12-13). Yet identifying stakeholders is not only done at
the start of a project, but rather, follows the twists and turns of the project and program itself.
And as Guzman and Feria (2002) and Boser (2002) point out, agencies and funders in these complex authorizing environments do not just accept the imposed expectations concerning evaluation. They negotiate the expectations, the resources and the processes of the evaluation work, just as they negotiate the work to be undertaken in delivering programs and services. The inherent contradictions of having multiple funders with diverse expectations makes this inevitable, even if it remains challenging and is usually undertaken with a great deal less power on the part of the delivery agents. This power differential is a key element of the way that complexity is relevant to evaluation in small nonprofit and grassroots organizations.

In discussing the issue of resources available for evaluation work, Fredericks et al. (2002) suggest that “(a)t a minimum, evaluators are counseled to avoid taking on projects where sufficient resources are lacking” (p. 12). And yet most of the organizations I work with do not have a choice about whether or not they do the evaluation. As a consultant, I do have a choice – I could hold out for more money, which would never happen – or I could pass on the project and let them find a volunteer, or do it internally, or find a less expensive consultant. The organization has to find some way to undertake the evaluation work, and almost inevitably, the resources available fall far short of the mark for doing the kind of evaluation that the funder and (usually) the agency want to do.

Not only do the agencies need to do evaluation, but the emphasis by administrators is on more traditional approaches to evaluation that focus on goal-oriented programs, with clear mandates and easily measurable outcomes. Part of the role that I undertake when working with nonprofit organizations is providing the knowledge and experience to recognize the forms of evaluation relevant to a project, and to assist in the negotiation of reasonable
expectations, particularly given the complexity of an environment involving multiple funders and assorted community stakeholders. While I may not be involved in the direct negotiation of the expectations at the start of the project – usually the consultant is brought in only after the contract is in place – I can often work with the funders’ representatives to negotiate evaluation expectations that more closely match the nature of the project. Sometimes, as in the example described in Vignette 5.1.2, the negotiation that I undertake addresses the perceptions of my client, which often go to the worst possible scenario in which the nonprofit agency anticipates that the focus is primarily or solely on accountability issues. Sometimes my efforts at negotiation end up focusing on both client and funder, as both have expectations about the possibilities for a range of findings that are not supportable given the program’s size, short duration, and limited resources.

Patton’s developmental evaluation approach (1994) seems increasingly relevant to the complex contexts found in grassroots and small nonprofit projects. Patton’s more recent work in this area (Campbell et al., 2003; Westley et al., 2006) has expanded to examine the ramifications of complex program settings. Developmental evaluation’s emphasis is on the emergent nature of programs, and “(p)aradoxically, openness to what emerges becomes the goal, even though one is simultaneously moving forward toward an envisioned future” (Westley et al., 2006, p. 238n). While the approach makes sense for the settings, explaining and negotiating such a picture of evaluation with clients and funders can be a challenge. Yet the benefits can be quite tangible. Abma (2000a, p. 461) describes how the search for goals can be part of the developmental process of a program in situations in which ambiguity is inherent in the context, or part of the rationale for the implementation. Vignette 5.5.3 describes an example in which the project coordinator and evaluation team allowed a high
level of ambiguity within projects, as a way of exploring the range of possible solutions to questions about what works in local program implementation.

**VIGNETTE 5.5.3 (NONPROFIT) – REFRAMING GOALS & ACTIVITIES**

I am meeting with representatives from each of eight communities participating in a health promotion demonstration project. Each person is taking a few minutes to fill everyone in on the status of his or her project. Martha, the representative of the most northern and isolated community, describes with power and enthusiasm her committee’s work in establishing what she calls a crisis line, and how it seems to have struck a chord with community members who have been using it far beyond expectations. The project coordinator, myself, another facilitator and several other community representatives glance at one another, somewhat puzzled – the community already has a crisis line available, and in fact it is an expensive program that may be in danger because it is rarely used by community members. We are wondering why they seem to be duplicating a service in a time of such fiscal restraint.

As people ask questions about the crisis line, and we develop a better appreciation of what the community is doing, we reflect on what Martha refers to as a crisis line, and reframe it as a peer-helping network that engages people across the community in helping themselves and one another. Further discussion and comparison illuminates a problem with the original crisis line – people in this primarily First Nations community are not inclined to call anonymous individuals based in another community. They want to talk with someone who knows who they are, and something about their history; someone who can appreciate the nuances of their lives and challenges, both culturally and personally. What makes the crisis
line work well in larger centres – its anonymity – seems to make it less suitable for this geographic location and for this population.

**REFLECTION 5.5.3**

Part of the challenge of using a developmental evaluation model with grassroots and small nonprofit organizations is that funders are reluctant to give up their project goal definitions. Such goals simplify and clarify accountability expectations for the funder, but they also do something more. They delimit the boundaries of what are considered to be appropriate solutions for the issues that lead to the development of programs. The strength of developmental evaluation – that it enables social innovation by expanding the vision of the evaluation beyond desired outcomes and hoped for results to encompass unanticipated consequences and tangential learning – becomes a deficit in the eyes of administrators who remain cognizant of the potential moral and fiscal imperatives of discovering alternative solutions or program possibilities beyond their immediate control. In such situations, funders must shift their focus to justifying inaction, or finding ways to limit the potentially significant obligations implicit in public discussion of options identified in evaluation. This is one reason why funders so closely guard not only the program and goal definitions used to frame evaluation, but the specific range of questions that are seen to be an appropriate mandate for the examination. Westley et al. (2006) argue that “(t)he narrow goal orientation of much traditional evaluation can result in missing … important unanticipated outcomes (p. 238n).” But such a goal orientation also serves to limit the culpability of funders from broader expectations on the part of other stakeholders and the broader public.

From the perspective of those in the vignette, allowing the project and the evaluation to proceed with an openness to emergent program goals and findings offers risks to all parties
in the setting. While the locally defined and implemented peer helping solution appeared to be effective in the short term, it had characteristics that made it problematic. It was demanding on local resources, and involved intensive personal investment by local actors – primarily volunteers – that would likely be difficult to maintain in the long term. It cast a specific funding source currently available to the community – the existing crisis line – as ineffective, and thus raised the potential danger of having it cancelled by the funder, with no guarantee that the funding would be transferred into some other community initiative. The short-term success of the peer helping effort shifted attention to the brief duration of the project, and gave an example to those within this and other communities that the funder seems to know what will work, but may not be willing to provide the resources necessary to make it happen. It sets up the implicit obligation to continue with what works, even when it is not clear how and why it actually did work, or what would ensure its continued success.

Developmental evaluation’s emphasis on looking cross-scale, allowing micro to macro reflection on programs and their contexts (Westley, et al., 2006, p. 239n), could potentially reduce funders’ control over problem definition, and empower those within communities to frame problems and solutions systemically. While many would applaud such empowerment, it remains problematic for those interested in maintaining control over resource expenditures. Returning to our previous discussion of the complexity inherent in having multiple funders and deliverers, having many stakeholders complicates the delivery and evaluation settings in numerous ways. It introduces more parties to any negotiation, to discussions about program goals, targets, delivery approaches and criteria for success. It introduces a tendency to search for compromise, potentially obscuring who is asking for what. And it shifts the programming emphasis away from risk, from the cutting edge of
ideas, and towards the safe implementation – something “we can all agree on” (Kohn, 1992, p. 130). In the previous vignette, the project coordination and evaluation teams experienced, or at least perceived pressure to reign in the possibly wayward project team who were introducing a new idea that seemed on the surface to be both counter-intuitive and potentially duplicative of existing services. Our ability to let it unfold and to see how it came together was not the typical experience for a nonprofit group trying something daring.

This dialectic between innovation and containment – between openness to emergent opportunities and following clearly defined and delimited goals and requirements – remains implicated in the contracting process, and the ways that funders delineate the boundaries of criteria for obtaining funding, and for demonstrating suitability for maintaining funding. Phillips and Levasseur (2004), describing a situation in which they have interviewed public servants and contractors concerning their experiences with contract development, noted that:

… government is focused on the “nuts and bolts,” projects that can be conceptualized, completed and evaluated, usually in the space of a fiscal year. So, the applicants respond accordingly: as the executive director of a multicultural organization working mainly on preventative issues told us, “I have learned how to take a preventative, proactive program and make it seem reactive.” For the most part, however, organizations were not simply learning to repackage their projects; they felt they were submitting more conservative proposals. (p. 459)

Disguising the nature of a project has been discussed by a variety of writers, but such a process inevitably is challenged by more traditional approaches to evaluation. Establishing flexible criteria for determining project success can alleviate some of the tensions for the project delivery team, but as the next vignette portrays, such a process needs to be ongoing.

**VIGNETTE 5.5.4 (GRASSROOTS) – Reframing Success & Failure**

It is near the end of the eight-community demonstration project, and I have been in meetings with the coordinator, Sophie, and agency director, Maggie, about the project and
the final write up of the community section of the evaluation report. The project has
undertaken a few twists and turns over the course of the year, and the implementation process
has differed from the initial plan. Sophie left to take another job a month prior to the project’s
completion, but has come back to tie up a few details of her report, and talk to me about the
project while I am back in the community for my final site visit. Her role has been taken over
by Maggie until the final report is submitted, and the director has arranged for me to meet
with a local group of volunteers, board members, staff and clients.

As I arrive, the coordinator and director are having a ‘spirited’ discussion about the
lack of congruence between the initial proposal and logic model prepared a year ago, and the
final description of accomplishments in the coordinator’s report. Maggie seems concerned
that “so much of what was supposed to happen did not happen,” and is questioning Sophie
about why the work wasn’t done. Sophie is becoming defensive in describing all of the many
unscripted activities that were undertaken, “because it turned out that we could not do the
community implementation part of the project until we had done the educational piece – we
didn’t have the mandate or the buy-in from other agencies to do what we wanted, so we had
to build that, or it would have crashed and burned.” It appears that from Maggie’s
perspective, the project has already crashed and burned, and so she is expressing concern that
in comparison to other projects, this one won’t look very good. I find myself working to
reassure Maggie that the project has not been a disaster, but she seems sceptical.

As we move to the larger meeting room for the community discussion, I learn that my
presentation to the community is the main topic. Since I had not known that the meeting was
taking place, and had no presentation prepared, this comes as a surprise. An even bigger
surprise is my discovery that the room is filled with almost forty people, most of whom have
concerned looks on their faces. Maggie introduces me as the evaluation coordinator for the demonstration projects, and suggests that perhaps I will be able to help them to “see beyond the failures of the project.”

In my brief thoughts about how I would participate in the meeting after I heard about it earlier in the day, I determined that I would start with a brief comparison of formative and summative evaluation, and how our emphasis is on learning. This still seems like a good place to start, although I quickly realize that I need to talk about this in a very practical way for the people in this room. Maggie’s portrayal of the project, and their own prior discussions (about which I know little), have left them apparently somewhat disconsolate. Luckily, I have an overhead transparency that compares all of the eight projects across a number of dimensions – including what they had set out to do a year ago. As I project the slide on a wall, I start to tell stories about each project – what they planned to do, what they actually did, and how that differed from their initial expectations. For each community, I am able to identify examples of how what they set out to do differed from where they ended up, and I then speak briefly to some of the learning that has come from this process. I do not refer to it as ‘unintended’ learning, but as learning ‘between the lines’ of the objectives of the various projects. I suggest that we are still in the process of identifying and compiling all the learning from the projects, and invite those present to share with me what they have learned, and their perceptions of the other demonstration site communities. The discussion becomes increasingly animated and positive, and although funding for this project is ending, people start talking about next steps and how to build on community efforts to date.
REFLECTION 5.5.4

From the beginning, the project depicted in the vignette was described and promoted as a ‘demonstration project’ in which the main goal was to try out innovative approaches to designing and delivering community-based health promotion strategies. The definition of ‘success’ in the initiative was clearly outlined as learning about the process of community level implementation of the strategies, and the key form of ‘failure’ envisioned was that we might lose the opportunity to share the learning from the projects with other communities. And yet even here we found communities building their own definitions and fears of failure into their self-evaluation – they needed no direction from the funder to frame the implementation challenges in this way. Indeed, my efforts in the situation focused on helping various stakeholders to look to the emergent characteristics of the situation, that the challenges faced and not surmounted over the course of the implementation were important parts of our learning that helped us to understand what does not work, as well as strategies we could employ in our efforts to make it work better in other communities.

Developmental evaluation’s use of complexity theory to direct the focus of evaluation on ‘side effects,’ ‘secondary effects,’ and ‘unintended consequences’ (Westley et al., 2006, p. 238n) can help to reframe success and failure, although community level program delivery staff and volunteers can find it challenging to break out of the accountability-oriented vision of evaluation that they have typically experienced, and will likely see again. They may reach the point of appreciating and even embracing it, but they must also be prepared to shift frameworks based on the approach endorsed by the next funder they work with. As the mediating consultant, I can provide a rationale for alternative approaches, and support them, but even when we have been clear about the approach, I can still be caught with the last
minute “but where are the numbers?” from either my community client or the funder. The more stakeholders involved in a discussion, the more likely it is that such questions emerge.

In their case studies, Phillips and Levasseur (2004) describe how funders’ accountability focus is on more than just evaluation – it also encompasses project audit, which can be unreceptive to changes or innovations within projects.

There was also a sense that innovative problem-solving during the course of a project was not particularly welcome, because a new innovation might not pass an audit if it had not been in the original proposal. Therefore, when innovation does take place, it has to be disguised rather than shared. (p. 459-460)

Disguising a project can happen in the proposal development process, during project implementation, or in reporting about both what the project did and what it achieved. In the vignette, if the coordination and evaluation team had demanded a stronger accountability focus, there would have been a great deal of pressure within the project to disguise the activities in the community and find a more positive way to portray activities and accomplishments. The likelihood of such ‘gaming’ activities (Smith, 1995) can be increased when there is a disconnect within the funder’s administrative process – when the individuals reading and judging the proposals are not the ones overseeing the project, or in some cases, reading the final project and evaluation reports. This represents another form of complexity introduced into the administrative process.

And yet the misrepresentation of agencies and activities is not only confined to the specific work of a project. Tassie et al. (1998) discuss a Canadian case study of how three funding organizations evaluated the same two agencies. The case study examined how different funders employed contradictory evaluation frameworks. The executive director of the agency dealt with this by becoming a “juggler” and possibly manipulating funder perceptions. In one instance and for one funder the agency depicted itself as a large multi-
program agency capable of integrating numerous programs and sophisticated sources of
funding, thereby achieving economies of scale, which made it attractive to the funder. To
another funder the agency could depict itself as a set of local offices with local knowledge,
capable of responding responsively to local conditions. In other words the agency could
depict itself as an established, well-connected and stable member of the social services
system, capable of providing a range of services, or if required, it could depict itself as a
community-based agency that represents local knowledge and responsiveness. Such a tension
is enabled and made more likely by the need to develop relationships with a variety of
funders and community delivery partners.

Such efforts to control the portrayal of agencies and programs represent an
understandable form of resistance to the contextual power of funders that nonprofit and
grassroots organizations experience. The implications for other agencies and clients can be
profound. As agencies attempt to fashion a public persona of competence and
responsiveness, they are balancing the need to appear bigger than they are – but not too big;
capable of delivering what the funder wants – but with responsiveness and understanding of
local conditions; and integrated into a network of local delivers with which they can partner
appropriately – or successfully compete with in garnering new contracts. And just as funders
work to lever additional or complementary funding from other levels of government,
agencies work to lever additional resources from other funders by demonstrating the
legitimacy that comes with having others identify the agency as worthy of funding. The work
and reports of evaluators potentially become another facet of this leverage – demonstrating
innovation, competence, thoroughness or sophistication of approach, depending on the need.
In this way agencies can appreciate the benefits of broader approaches such as developmental
or goal-free evaluation, but they also depend on being able to demonstrate a willingness and capability of undertaking and even embracing more traditional evaluation, which situates them closer to the mainstream.

Finally, there are several implications of the complexity inherent in having serial and concurrent multiple funding of projects and agencies. One of the unintended consequences of not funding repeat projects is that agencies come up with many ‘versions’ of programs that each address their needs in different ways. Whether or not they have employed an evaluation as part of the project, agencies naturally self-evaluate in that the subtle changes they make in adapting programs as they submit proposals for new projects that represent something new and innovative to project funders, do represent a cumulation of local knowledge and experience about what works and what does not. New proposals address gaps and changes within the community delivery network. They can represent a growing understanding of the issue and program based on client feedback, stakeholder input and concerted lobbying by interested parties. They can reflect new perceived opportunities or represent a fall back position to a more basic approach. But the process also reflects the tension between innovation and conservatism. The innovations are required, but they should be incremental rather than bold (Kohn, 1992, p. 130). They should be measurable rather than exploratory. And they should be achievable within a year and become ‘self-sustainable’ after that.

I have seen many agencies adapt programs over an extended period of time, with some of these iterations ‘evaluated’ in some way, but rarely is the long-term process documented in any detail. Such adaptation processes represent wonderful opportunities for long-term meta-analysis of programming efforts. But in my experience such efforts are rarely undertaken, in part because funders treat each iteration of a program as a self-contained and
unique effort. And project proposals tend to enable such fractionation of the issues, as they seek to provide a balance: building on the experience from which the new project has grown, while simultaneously distancing the new project from predecessors to make it appear ‘new.’ The end result that I see is a social program landscape littered with the debris of innumerable fascinating and worthwhile efforts, with both the ideas and their impacts lost to history and to our collective learning.

This section has examined a variety of facets of the complexity inherent in the delivery and evaluation of programming in grassroots and small nonprofit agencies. Multiple funders, multiple programs, and multiple delivers in a complex web challenge all parties to understand the shifting landscape of allegiances, connections and priorities of stakeholders. While developmental evaluation appears to hold promise in that it acknowledges the inherent complexity of these contexts, its focus beyond the boundaries of the narrow concerns of traditional evaluation – essentially what makes it valuable as a method of reframing and exploring innovation and unexplored solutions – can make it risky for many stakeholders. It has the capability of opening Pandora’s box, which makes it dangerous and potentially revolutionary in its implications for future programming and funding.

Summary

Chapter Five examines evaluation contexts from the perspective of the impacts that devolution has had on grassroots and nonprofit organizations. I reflect on five themes arising from the process of devolution, examining some of these impacts on program delivery in the sector, but emphasizing what the devolution context means for the practice of evaluation.

The five broad themes described in Chapter Five crosscut in diverse ways. At the heart of the trends are stresses experienced by grassroots and small nonprofit organizations
growing out of the increased accountability demands placed upon them by participating in
the contracting process, and challenges to the capacity of the organizations to take on the
work of devolved delivery and evaluation. As opportunities to obtain operating grants
decline, organizations enter into contracting relationships with funders, who introduce
enhanced accountability expectations. The organizations become locked into an unrelenting
focus on resource stabilization, which ironically means that they end up taking on new
contracts to maintain their ability to undertake prior contracts. Capacity stresses are reflected
in staff and volunteer turnover, burnout, stress and the increasing use of contingent labour
and resources – people are working ‘off the sides of their desks.’ This leads to pressures to
adapt and shift the work undertaken – creaming those clients most likely to succeed, thus
ensuring that the organization meets seemingly arbitrary targets and maintains contracts –
and generally shifting the priorities and mandates of programs to suit funders’ needs rather
than those of the local communities and clients. The competitive contracting process has the
capability of changing relationships of agencies within communities as they compete for
funding or clients, and introduces pressure to establish linkages beyond the community
through federations, or alternatively shift the competition to one between communities, to
programs within an agency, or clients within a program. Concurrent demands for agencies to
partner with other delivery organizations, or lever contingent funding from other funding
sources introduce further complexities to the contracting situation, with overlapping
mandates, timeframes and funding periods, and competing accountability expectations, all
undertaken in a context of dramatically increased resource stresses exacerbated by turnover
and burnout. Contracts demand innovation while discouraging risk; they require programs to
demonstrate demanding outcomes while providing inadequate resources in either time or money to successfully achieve those outcomes, and typically few resources to measure them.

This context is experienced by grassroots and nonprofit organizations as a succession of overlapping tensions: accountability vs. autonomy, stability vs. growth and change, solid service provision vs. innovation, competition vs. partnership, funder priorities vs. local responsiveness, short-term and simplified solutions vs. complex and long-term problems, and standardized services vs. client-centred ones. Not surprisingly, nonprofit organizations can develop relationships with funders based on control and resistance to that imposed control – looking for ways to maintain some autonomy and lever resources rather than having mandates determined from without. One arena in which this battle for resources is played out is evaluation, a key new component of the accountability dance.

These tensions provide both opportunities and tradeoffs for nonprofit organizations. The challenges to maintaining local connection with other delivery organizations and clients can be offset by the potential wealth of new partner organizations, ideas and opportunities through the outward focus of linking beyond the community. They can also gain a voice at a larger table, one that may have been inaccessible locally. The benefits to growth trade off with significant challenges – long-term commitments to staff, program delivery, and getting on the proposal and resource development merry-go-round. On the evaluation front, engaging in more clearly defined program definitions and logics trade off increased program clarity and continuity with a tendency towards less flexible, innovative and emergent approaches to program development and delivery.

What does all this mean for evaluators? One tension that I experience with grassroots and nonprofit organizations in the context of devolved programming is their decision to take
on evaluation work in-house or through the use of a consultant. This is a significant decision for many organizations that are already stressed for resources. They rarely have someone internal to the organization who can take on the responsibility for evaluation in anything beyond a very cursory way. This is one reason for the strong emphasis in ‘how to’ resources for nonprofit evaluation that try to make distinctions between simple output measurement and the identification of program outcomes. The most traditional and basic approaches to evaluation are the ones most easily shared with in-house nonprofit representatives who have little evaluation experience.

On the other hand, working with a consultant represents both potential benefit and risk for nonprofit organizations. Consultants are typically more expensive than small organizations are prepared to work with; they may offer sliding scales for different types of client, but it is an expenditure that trades off with scarce financial resources that can equate to a great deal of program implementation opportunity. Consultants also are outsiders who become aware of the organization’s internal workings, which for small agencies can appear chaotic, on the brink of disaster, and contrary to the professional public face the organization is trying to develop and portray. This is a risk for the organization, because the consultant has allegiances to reporting what happens at the agency to the broader community and to funders, even when the agency is the client for the evaluation work. On the other hand, the consultant brings a variety of resources, adds legitimacy to the evaluation process, and can provide the agency with a way to demonstrate to present and future funders competence, ambition, and openness to transparency, no matter what the findings of the evaluation might be. Yet if the consultant is not familiar with the contingencies of program implementation in a context of devolution, the organization runs the risk of being judged by standards and tools
inappropriate to the task and setting. In my experience as a consultant transplanted from conducting much larger scale evaluation of provincial and federal level programming, agencies sometimes cope with this risk by undertaking a program of ‘consultant education’ about the unique requirements of program delivery and life in a nonprofit organization.

The external evaluator working with devolved program delivery in grassroots and nonprofit organizations must also cope with a variety of tensions inherent to the context. A key one is between perception and actuality – the communication between funders and deliverers is fraught with misunderstandings and haphazard interpretations. This is particularly apparent in smaller nonprofit and grassroots organizations that are trying to appear larger and more experienced than they are in order to compete for funding, and that are subsequently reluctant to ask clarifying questions that might demonstrate their level of inexperience or ignorance about the rules of the game.

Besides clarifying processes for clients, evaluators can play a mediating role between funders’ evaluation expectations and clients’ efforts. This can involve interpreting, facilitating, and negotiating expectations, requirements and processes. Section three of the next chapter will explore these ideas in more detail.

Because of the long-term relationships that they can develop with organizations, particularly in view of the turnover of staff and volunteers in agencies, evaluators can come to represent the corporate memories of these organizations, or through their documenting efforts, their reports with thick descriptions of programs and their rationales can provide that function. Further, because of their involvement with several programs in an agency, sometimes over time, and possibly across several agencies in a community, evaluators can be well placed to observe mandate drift, agency changes, and even shifts in the diversity of
programming at the community level. As such, they can play an integral role in documenting and articulating meta-analyses of programming issues. They may also have opportunities to bring such analyses and issues to the forefront of discussions and dialogue with funders and other communities. Section four of Chapter Six further explores these ideas.

Part of such efforts would encompass seeing the shifts in transparency expectations from government down to local communities. As local communities and community based grassroots and nonprofit organizations take on some of the evaluation tasks of devolved delivery, these tasks become less open to public scrutiny in that the findings of such evaluation work are not necessarily placed into a public forum. This represents a reversal of the trend that has seen most federal and provincial level evaluation reports made available to the public either on request or increasingly, on-line. The high number of local evaluation reports makes efforts to introduce more transparency into this level of reporting challenging. Their development in a context of local community delivery, making it more problematic to maintain the confidentiality of clients and deliverers within agencies, introduces further risk. When consultants work with local agencies to either conduct or facilitate such evaluation work, part of their focus must of necessity be on analyzing privacy threats of reporting, and working with stakeholders at several levels to ensure safeguards are in place to protect information, and to educate participants in the processes. This is particularly crucial in settings in which more participatory approaches are used, as those conducting evaluation need to learn of their own potential complicity concerning some of the threats to privacy inherent in doing evaluation. The second section of the next chapter explores some of the implications of participatory approaches to evaluation in a context of devolution.
The first section of the next chapter examines theory-based evaluation – a major trend within evaluation over the past two decades, and one of great import for nonprofit and grassroots organizations because of the growing expectations that logic models – a common form of representation of program theories – have become a strongly endorsed tool demanded by funders and administrators working with delivery agencies in the sector.

The next chapter examines four themes that reflect trends within the evaluation literature. My reflection on the implications of devolved delivery contexts for broader trends in the evaluation field will build on the lessons learned and observations through the vignettes in Chapter Five. My goal in Chapter Six is to examine the themes and their relevance and applicability to conducting evaluation in a context of devolution. Devolved program and service delivery implicates evaluators’ diverse roles, inclusive approaches, and role in fostering dialogue. Evaluation conducted at the margins challenges traditional approaches to evaluation. In so doing, it raises questions about those traditions as well.
CHAPTER SIX:
CASE STUDY ANALYSIS: EVALUATION THEMES

Evaluation within the ideology of the New Public Management is focused on observable, often quantified, hopefully objective indicators of what difference programs are making in the lives of participants. And it is focused on grounding decisions in scientific evidence. These are worthy and important aims. But, they are limited ones, precisely because they are constructed from the ostensible safety of the air bubble. They are aims that lack democratic engagement and thus political potency.


Evaluation is experiencing a time of dramatic growth and change. The discipline has been expanding to new contexts and challenges, and has grown in its scope, breadth and range of influence (Stame, 2006; House, 2004; Alkin, 2004; Khakee, 2003). I have chosen four broad themes apparent in the recent literature that are particularly relevant to understanding evaluation in a devolved program context. My goal is not to reframe or reconceptualize evaluation through these themes, which would be beyond the scope of this study, but to identify how the themes both represent and are implicated in the changes demanded by the context of devolved program delivery and evaluation.

This chapter reflects on the relevance of these themes for the devolution context. The broad themes are: 1) theory-based evaluation, 2) inclusiveness, 3) evaluator roles, and 4) dialogue, deliberation and democracy. The themes are crosscutting, and some examples and vignettes are appropriate for understanding multiple themes. Each section provides three types of information: 1) examples from the case studies in the form of two or more vignettes – at least one representing a small to medium-sized nonprofit and another representing a grassroots organization, 2) relevant concepts and examples from the literature, often using case studies conducted by others, and 3) reflection on the vignettes and concepts representing an analysis of the relevance of these ideas for evaluation practice and theory.
Theme Six: Theory-Based Evaluation

This section explores the recent emphasis on theory-based evaluation, and its relevance for nonprofit and grassroots organizations. While the idea of using a program’s background assumptions and ‘theory’ to guide evaluation work has been discussed in the literature for over thirty years (Rogers, et al., 2000), it did not become a significant movement within the discipline until the late 1980s and early 1990s (Worthen, 1996, p. 169). Chen and Rossi (1980, 1987, 1989) and Chen (1990, 1994b) described this as a ‘theory-driven’ approach to evaluation, while others (Worthen, 1996; Weiss, 2000) have used the more general phrase ‘theory-based’ evaluation. Theory-based evaluation was proposed to address the dominant methodologically driven and a-theoretical ‘black-box’ experimental emphasis on the relationship between program inputs and outputs then common in evaluation (Chen, 1990, 1994b; Worthen, 1996). A key consideration was the quest to understand the background assumptions and mechanisms through which programs achieve their effects – detailing the chain of activities connecting assumptions and rationale to program outcomes. Another important emphasis for the model was articulating this in terms of a set of causal processes, and providing the opportunity to allow evaluators to discriminate between theory failure and implementation failure (Bickman, 1989, p. 387; Chen, 1990, p. 29).

The theory-driven approach, particularly as promoted by Chen and Rossi (1980, 1987, 1989) was both welcomed, and received with some caution. Perhaps the main benefit that Chen and Rossi anticipated was addressing the issue of goal definition in summative evaluation. Program theories represented a means of bypassing battles over identifying goals by linking context and program assumptions with activities, and developing a more comprehensive and nuanced articulation of multiple goals relevant to evaluation. Beyond
this, it also allowed the evaluation to extend beyond a simple consideration of goals – either ‘official’ or operative – to encompass unintended program effects, and provide a basis upon which funders, administrators and deliverers could find common ground and evaluators could experience less resistance to their efforts (Chen, 1994b; Worthen, 1996).

Chen (1990) defines program theory as:

... a specification of what must be done to achieve the desired goals, what other important impacts may also be anticipated, and how these goals and impacts would be generated. This definition implies that program theory consists of two parts. The first deals with what the structure of the program should be, including such things as treatments, outcomes, and implementation processes that are related to the values of the program; this part is related to prescriptive theory. The second part deals with what are the underlying causal mechanisms that link the relationships among program treatments, implementation processes, and outcomes; this part relates to descriptive theory. (p. 43, italics in original)

In this way, Chen focuses on not just a program’s goals and anticipated outcomes, but the processes through which those goals are to be achieved, the activities and presumed relationship between those activities and the goals and outcomes. The approach still clearly depends on identifying and articulating goals, but tries to frame them within a broader context. Indeed, in his book, Chen (1990, pp. 168-185) goes to great length to explore the nature of goals, and their place in framing the evaluation and building theory.

The initial critique of theory-driven evaluation encompassed a variety of concerns, and challenged not so much the value of using a theory-based approach, but the feasibility of doing so. Bickman (1989, p. 388) asks whether theory-based evaluation is affordable. He notes that because it is more comprehensive and encompasses more detailed planning and measurement development than traditional evaluation, it takes considerably more time and subsequently, is more costly than previous evaluation approaches. Further, Bickman notes that the majority of programs do not have well articulated theories, and that this requires more time to consult with stakeholders and construct theories at the front end of evaluations.
Indeed, this is a concern expressed by Chen and Rossi (1989), who note that theory-driven evaluation that deals with multiple issues “… is expensive and time consuming. … comprehensive theory-driven evaluations are [likely] to move at glacial speeds” (p. 300).

Another facet of the critique of theory-based evaluation grew out of this view that it demands not only additional cost and time, but also new skills and knowledge. Patton (1989) argues that:

…use of the model requires a level of research sophistication that will be lacking in most stakeholders resulting in the evaluation conceptualization being controlled by sophisticated researchers who understand the model with the consequence that nonresearch stakeholders are left out of the process. (p. 377)

Patton’s concern here is not just with the exclusion of stakeholders, but in the demands made upon evaluators, who are taking on additional responsibilities under the approach. Bickman (1989) also notes this new knowledge requirement, and points out that it involves two types of knowledge – one concerned with how to develop and articulate program theories, and the other being a much more comprehensive and detailed content knowledge about the program area – “more than passing familiarity with the program content” (p. 389). For Patton, the latter is obtained primarily through consultations with stakeholders. Chen and Rossi (1989) respond to this idea in a way closer to Bickman’s position, arguing that the evaluator must include stakeholder views in program theory development, but they must also go beyond stakeholders’ perspectives on the program, and use their knowledge to build broader and more general theoretical understanding (p. 305). It is clear that Chen and Rossi are in basic agreement with both critiques of their theory-driven approach – that it is more costly, time consuming, and demands new skills and new knowledge of evaluators and stakeholders.

A more significant critique of the theory-driven approach to evaluation reflects its potential breadth of applicability to the diversity of evaluation contexts. For Patton (1989),
the key challenge of the approach is that it is too narrow – particularly in its apparent focus aimed at generalizing causal linkages. He argues “program evaluations which focus on generating generalizations about causal linkages are a small and narrowly conceptualized subset of all the activity in which evaluators engage” (p. 375, italics in original). Here Patton is concerned with the range of questions and issues of import to stakeholders, which encompass much more than trying to understand either the achievement of intended outcomes, or the accuracy of the underlying theories relevant to how the program performs. His emphasis is a practical one, concerned with evaluation as it is conducted in the field, but also reflecting a more general observation that the theory-driven approach seems to reflect an academic and ‘esoteric’ emphasis out of place in the real world of evaluation practice.

Smith (1994) addresses and builds on similar concerns, arguing that the theory-based approach is relevant to only certain types of programs, particularly “those with an explicit, well-defined intervention or treatment” (p. 84). As such, most programs that are evaluated are not sufficiently well established in terms of their “purity, specificity, dosage, intensity, and duration” (p. 84, referencing Scott & Sechrest, 1989). Smith further notes that many – perhaps the majority – of programs do not have their causal mechanisms well documented.

Smith’s key concern is that the reality of implementation for many programs does not meet the strict and exacting standards required of evaluation as envisioned by those promoting the theory-based approach. Programs are typically less bounded and stable than expected in the approach, reflecting constant change and adaptation, even volatility, and this emergent nature of programs may represent not problems but important strengths:

Program implementation may vary dramatically across sites, across time, and across participants. This variation may reflect a design strength (wide applicability and adaptability), rather than a failure of design clarity and managerial control. In such cases, program theory should provide a descriptive view of the program as seen by
different stakeholders—seen by some as a tightly defined intervention, by others as a dynamic system in flux, and by still others as developmental and emergent. Program theory should thus reflect the program as it is, rather than as it should be according to some preset definition of program as technical intervention. Such variation in the program’s theory certainly makes the evaluator's job more difficult, but programs should be constructed to be effective, not to be easily evaluated. (pp. 85-86)

Underlying Smith’s discussion is an appreciation that many programs – particularly new, emergent and dynamic programs such as those typically found in grassroots and nonprofit organizations – do not have a clearly articulated program theory. Further, Smith argues that the complexity and ambiguity within these situations suggests a need for multiple working program theories that take into account the varieties of program context (1994, p. 86), an argument anticipating the more in depth treatment later provided by Weiss (2000).

The concern about whether programs have clear or articulated theories appears to overlap with the other initial critiques of theory-based evaluation – the potential costs in time and resources incurred in developing theories, the technical skills and substantive knowledge needed for developing program theories, and the narrow emphasis on programs for which theories are either available or suitable. Patton (1989) questions the apparent assumption by Chen and Rossi “that programs are undergirded by theory” and argues “that clearly conceptualized programs are a rarity” (p. 376). Even Chen and Rossi (1989) lament “the embarrassing absence of theory relevant to most social programs” (p. 300), and see it as the main barrier to having theory-based evaluation more widely adopted. Chen and Rossi, Patton and others suggest that theory-based evaluation could and should encompass program theory development, and yet the initial vision was that this was a complex, time-consuming and resource intensive process, one requiring the hands-on skills of knowledgeable evaluators, or at least their substantial guidance. Patton (1989, 1997a) in particular, argues that working with stakeholders to articulate a program’s ‘theory of action’ represents an important task.
Over the decade between the late 1980s to the late 1990s, the growth of the now nearly ubiquitous logic model came to represent the primary mode of articulating a program’s theory. By the end of the decade, in a discussion about theory-based evaluation, Rogers, et al. (2000) were able to say “(m)ost program theories are summarized in a diagram showing a causal chain” (p. 7). While logic models had been around and in use for at least twenty years, the decade saw an explosion of use, particularly as the approach became popularized by such organizations as the United Way and the W. K. Kellogg Foundation (2004) through freely available guides (McLaughlin & Jordan, 1999, pp. 65-66). McLaughlin and Jordan (1999) talk explicitly about how logic models can describe expected performance, as well as how they articulate the theory behind a program. They quote Chen (1990), Patton (1997a) and Bickman (1987) and discuss how the program’s theory of action and stakeholder perceptions of a program may be represented using logic models. Numerous others have also made the explicit link to how logic models can be used to represent program theory. For example, Savaya and Waysman (2005) argue that program theory is widely recognized as important for both developing and evaluating programs, and that since developing program theory can be a complex and challenging task, the logic model represents a simple, powerful and “ideal tool” to guide practitioners in doing so (p. 87).

And yet how well do logic models describe and articulate program theory? Rogers, et al. (2000) appear to accept as unproblematic that logic models can represent program theory, and that they often do, particularly in nonprofit organizations using the United Way approach (p. 6). However, elsewhere in the same volume, Rogers (2000) cautions that such logic and causal models are simplified representations of reality (pp. 48, 54-55). Similarly, another author of the article argues program theories tend to be far less detailed than they need to be.
in order to successfully replicate programs – they do not provide the level of detail needed to reflect the complexities and subtleties of the programs, or of their contexts (Hacsi, 2000). Indeed, it is the complexities of context that Rogers (2000) is making a case for in recommending new graphic forms of causal modelling. In the same volume, Funnell (2000) also proposes a ‘program theory matrix’ – a variant on the logic model intended to provide more comprehensive detail and allow for a more complete portrayal of program theory relevant to an evaluation.

Leeuw (2003) further argues that program logics rarely go beyond the level of theory articulated by stakeholders, and are typically incomplete constructions:

An important difference exists between program theory and program logic. Program logic often specifies the inputs and components of a program, as well as short-term and long-term outcomes, along with the assumed linkages among these. However, a program logic rarely outlines the underlying mechanisms that are presumed to be responsible for those linkages. (p. 6)

In this way we return to the concern expressed by Chen and Rossi (1989) with Patton’s emphasis on having stakeholders articulate the theory of action – that while there is a significant role for the evaluator in helping stakeholders articulate theory, the evaluator’s attention should be more broadly focused on helping to develop alternative representations of and modes for developing program theory (p. 305). Both Funnell and Leeuw offer ways for evaluators to help develop such comprehensive and alternative representations of theory.

McLaughlin (1999) and Funnell (2000) also note that logic models do not only represent program theory, but have come to constitute an important element of the expected accountability matrix for programs, and have become implicated in the performance measurement movement. Funnell, in particular, finds this to be problematic, in part because of the simplification represented by the typical logic model and demanded in performance
measurement approaches, noting that in this way performance indicators may be the antithesis of program theory. Section 5.1 in Chapter Five examined this relationship.

There is a significant irony in the way that simplified logic models have come to represent program theory in much evaluation. Chen (1990, 1994b) argues that theory-driven evaluation is warranted because “method-driven evaluations simplify complicated evaluation tasks” (1994b, pp. 229-230). So for Chen, incorporating program theory represents a way to introduce a suitable level of complexity into evaluation work. His approach articulates in detail how program theories can be constructed, should take into account broad context and unintended consequences of programs, and can lead to fuller and more textured discussion about the worth of programs and the problems and challenges of implementation.

A second irony represented by the history of theory-based evaluation lies in the original critiques by Patton (1989) and Smith (1994), among others. Their focus is on the explicit emphasis within theory-driven evaluation on causal explanation, and the implied reliance on experimental design approaches. While Chen (1990, 1994a) goes to some lengths to argue the merits of seeing theory-driven evaluation as encompassing both qualitative and quantitative design approaches, respectively implicated in program theory development and testing, Patton, Smith and others focus on the narrowness of the model as implying an experimental design model relevant to a small subset of programs. Rogers, et al. (2000) attempt to introduce a new phrase to represent theory-based evaluation – ‘program theory evaluation’ (PTE) – which they argue represents a deliberately broader definition, despite its focus on only one of the six types of theory-driven evaluation articulated in detail by Chen (1990); the one type that happens to most closely represent the narrow, experimentally relevant picture of theory-driven evaluation that was the focus of critiques by Patton and
Smith. This irony is further belied by the way that the majority of articles in the special issue refer to ‘theory-based’ evaluation to discuss other diverse facets of program theory in use.

Chen’s concern with the role of evaluation in demystifying programs for stakeholders (1994a, p. 81) encompasses not just testing programs but also helping to develop program theory or make it explicit (p. 82). He is clearly aware of and envisioning theory-based evaluation as a broad tool for stakeholders, including those involved with emerging programs such as those commonly found in nonprofit and grassroots organizations. In reflecting on the use and development of program theory in community-based programs, Leviton (1994) builds on such a vision. She examines programs within these contexts, and their relative ownership by the communities. Referencing McKnight (1991), Leviton examines the challenges in identifying and applying program theories when they are “outposts of institutions” (pp. 89-90) when it is not clear whether the program is reflecting local values and development, or those imposed from outside or by a funder. The bottom line for Leviton is that perhaps “evaluations conducted in community-based projects are better for hypothesis generation about services than for hypothesis testing” (p. 92).

And perhaps this is a key emphasis that we can apply from theory-based evaluation for thinking about evaluation in nonprofit and grassroots organizations: that while not all, or even most such programs have explicitly articulated theories, most are ‘undergirded’ by theory – or perhaps many theories. And these theories may not be well understood, even by practitioners. Evaluation can help to discover and articulate these theories, and to include them as part of the evaluation if possible or appropriate. Developing and articulating program theory can be challenging; doing so through the use of logic models represents the further challenge of using a tool that may be too blunt for the purpose, but also one that has been
incorporated into the toolkit of other stakeholders for different purposes, including funders who have made it part of the accountability dance.

The following vignette explores an example of how logic models may be applied in support of developing and articulating program theory in a nonprofit evaluation setting. It addresses how logic models can be very useful, even when they have significant limitations.

**VIGNETTE 6.1.1 (NONPROFIT) – BUILDING A LOGIC**

I have begun the task of evaluating the programming efforts of a medium-sized community nonprofit organization. The primary funder – a national department overseeing the work of hundreds of comparable organizations across the country – has asked for a complete review of the agency’s programs in its broad funding envelope, anticipated by the organization as a required accountability exercise every five years. As I start reviewing the agency’s many program areas, the director hands me the original terms of reference for the agency, outlining the set of criteria and expectations upon which programming was to be developed. The terms of reference was part of the original proposal to establish the current slate of program areas, and was forwarded to the director by the funder’s representative as the starting point for the evaluation. It is not clear that the agency has had a copy of the terms of reference on hand for the past several years. The document describes a set of program areas in broad terms with matching performance criteria that appear at first glance to be marginally relevant to the stated programming efforts. The director tells me that the programming was developed based on this document, and that these criteria are to be the basis for the evaluation.

As I begin to examine the agency’s documents and databases, and observe the programs and services offered in the agency’s core and satellite offices – most of which fall
within the domain of the department’s funding envelope – it becomes apparent that the current programming bears little resemblance to the original template described in the terms of reference. When I discuss this with the director, I am told that the criteria and expectations as outlined in the terms of reference must be the focus for my efforts – and that I need to document the agency’s success in achieving improvements relative to the original baseline community survey conducted at that time. As I talk to the director and try to build a mental map of the programming areas, activities of agency staff and services to the community, I am increasingly struck by how significantly it differs from what was initially anticipated in the terms of reference.

Despite the clear direction and expectations of the director, I focus my initial efforts on detailed interviews with staff in each of the offices, program participants and community partner organizations. Based on these efforts, and well into the official review process, I have developed a detailed model of how I perceive that the agency is structured and the range of programs in place. It has some similarities to the original model, but represents a significant departure from both the terms of reference and the picture described to me by the director. In anticipation of taking the next step of validating this detailed description with staff and community partners, I arrange to meet with the director and share my vision and interpretation of the agency. I have compiled the information in graphic form as a logic model encompassing five broad program areas, with links to sub-programs, offices, community partners and client groups. Based on my previous discussions with the director, I am anticipating that this meeting may represent a turning point for the evaluation, as I will be strongly suggesting that holding on to the original model described in the terms of reference...
is neither reasonable nor feasible. Not only would many of the data collection activities be relatively meaningless, they could only serve to portray the agency in a negative manner.

As we meet, I tell the director that I plan to frame the balance of my evaluation efforts in examining the programs and services as actually offered, and using the logic model as a point of discussion with staff and other stakeholders to validate what the agency does, what stakeholders feel the agency’s mandate should be, and the gaps in and effectiveness of the programming that does exist. The director lays the logic model – outlined on a large 11 x 17 inch sheet of paper – on the desk and peers at it with a look of intense concentration. I glance at my colleague, who has been helping with interviews and constructing the model, and we wait with increasing suspense and some anxiety as the director seems completely oblivious to our presence in the office. As her eyes grow wider and the look on her face becomes more serious and seemingly incredulous, we come to expect the worst. After about five minutes of intense scrutiny, the director suddenly exclaims, “This is fantastic – *I can use this!*”

Our ensuing discussion of the logic model clarifies several points of confusion about the agency’s programming, and the director fills us in on some of the history of changes to the programs that have occurred over the past few years. We follow up with the funding department’s representative, who is pleased to have an updated and clarified picture of the agency’s programs and services, even in draft form. Further discussions with community partners, staff and a comprehensive survey of community stakeholders provide additional detail about the diverse range of expectations about the agency – many of which have not been addressed by activities or programs to date.

The preliminary logic model serves to reframe the evaluation, and sparks discussion among community stakeholders about where the agency does and should fit into the broader
support network. The lack of clarity about agency programs is not resolved by the logic model, but gaps in understanding both within the agency and among community stakeholders help to point out where the next efforts need to take place, such as identifying the rationale behind the different health promotion activities that staff undertake, and the complementary and sometimes competing efforts by other staff, volunteers and community players.

**REFLECTION 6.1.1**

By initiating the discussion about program activities, even though my evaluation team will not be there to see it through, we are helping the agency to identify and clarify available and needed resources. The lack of a coherent set of program theories justifying the program and staff activities under review represents a challenge to the agency and to the evaluation effort. The new logic model serves to sketch the outlines of some of those program theories, but is not comprehensive enough to provide detail of the assumptions and rationale underlying the links between goals, activities and intended outcomes. The agency has developed programs based on perceived need, and drawn from diverse sources – including existing programs and services of smaller grassroots agencies that have been absorbed by the larger nonprofit organization – to construct a seemingly haphazard and at times incongruous range of activities. In many cases it is hard to identify groups of activities as a ‘program’ or ‘programming area’ – but our efforts to group them offer several interpretations and contrasts. They incorporate input from local stakeholders, as well as our knowledge of relevant programming and ideas from other agencies and communities. As part of my final report, I suggest a variety of resources that can help staff to construct such a rationale in more detail. These include both scholarly articles related to the substantive emphases of the
activities, and links to programs and services provided by similar organizations that have a clearer overall rationale for the diverse programming efforts they offer in their communities.

The evaluation effort described in Vignette 6.1.1 portrays a process of working to articulate agency activities, program goals, and some aspects of a theoretical backdrop to the programs and activities. This process involved investigating and negotiating stakeholder perspectives about what was done, why it was done, and what should be done by the agency, as well as incorporating information from other sources – including programming literature, the experience of the evaluation team, and reports and programs from other agencies. In terms of Chen’s model of theory-driven evaluation (1990, p. 43-44), the evaluation team was working to construct both normative and causative theory, the former relating to program goals and how the programs should work, and the latter relating to the actual processes and mechanisms through which outcomes are achieved.

Despite its emphasis on moving beyond the limitations of goal-oriented evaluation, Chen’s approach – particularly normative outcome evaluation (1990, p. 91) – still relies heavily on articulating program goals, even if that process is facilitated by evaluators. Chen (1990, p. 174) strives to improve the process of goal articulation through an intensive examination of types of goals, distinguishing between ‘official’ and ‘operational’ goals (referencing Perrow, 1961, p. 855), which are in effect very similar to ‘formal’ and ‘real’ goals, respectively (referencing Etzioni, 1960). This valuable discussion contrasts how vague and idealistic ‘official’ or ‘formal’ goals can provide the rationale that binds coalitions, while specific and practical ‘operative’ or ‘real’ goals that involve the details of resource allocations or value trade-offs can contribute to conflicts among stakeholders (p. 176). Such goals can be contradictory despite being considered accurate and useful for a program and for
different stakeholders, depending on the context and the use. The report prepared for the evaluation described in the vignette contrasted the diverse models and articulated visions of the agency’s programming provided by different stakeholder groups, and offered these as a tool for developing and facilitating a broader discussion within the agency and beyond. Rather than choosing one vision, the process and stage of development of the agency seemed to call for such a deliberative strategy, accommodating and representing the diversity of views, while enabling further discussion that could encompass goal clarification, but also development and negotiation about the nature of the theories that could encompass this.

Chen (1990) offers suggestions on how to identify plausible goals that are valued by stakeholders and also actually pursued through program activities (p. 182). In effect, his approach is to look for goals that actually have sufficient resources committed towards them, and that are consistent with current understanding about how the program can and should be addressed – in other words, prior theory, knowledge or experience about the causal linkages between goals, activities and outcomes (p. 183). Chen’s quest of providing these guidelines is noble, but in practice the approach replicates the problems of goal-oriented evaluation. The first criterion he suggests is an empirical question potentially answered through an evaluation, but hard to know prior to actually doing the research. The second criterion is open to the types of critique noted by Bickman (1989) and Patton (1989) – that the knowledge about the state of the art in relevant theory, knowledge and experience demanded may be beyond the scope of evaluators, at least prior to doing the evaluation work.

Friedman et al. (2006) offer an active process for resolving disputes about goals, defining and clarifying them based on evaluating the goals, and not just the programs, and using a questioning approach aimed at discovering why stakeholders hold to their respective
goal definitions (pp. 202-203). This approach tries to build upon a recognition of the different roles that goals have for various stakeholders, but it is ultimately unsatisfying in that, 1) it depends so heavily upon a process that brings stakeholders together, when it is often very challenging to either identify all stakeholders or bring them together, and 2) it assumes that stakeholders are relatively evenly matched in power or stake in a program.

Introducing a theory-based approach does not appear to be a panacea for the goals problem. It still encompasses many of the challenges of goal definition and clarification, and further, it may in itself constitute the same problem in a different form. Weiss (2000) argues that for any program, there may be many possible theories to choose from. It is usually possible to articulate more detail about the mechanisms underlying a program or extrapolate more detail about the chain of implementation steps involved. Some of these theories are normative, and developed from the views of stakeholders. Others are adopted from the academic literature. Normative theories tend to focus almost exclusively on the program’s elements, and not take into account alternative factors of context beyond the scope of the theory. Theories from the literature may be a poor fit with the specific context, too abstract, inappropriately adapted, or be out of date with respect to current thinking in the field. In short, there are no guarantees that theories represent an accurate fit with programs as they exist, or that they will remain uncontested.

Shaw and Crompton (2003) further argue that “in choosing a theory we also exclude other theories and exclude areas of information that lie outside of the particular theoretical framework of choice” (p. 193). In doing this, the evaluator narrows the definition of the problem, and narrows the range of solutions considered appropriate to that problem. Program theories tend to frame programs in terms of a limited vision of the problem, and frame
solutions with a view to the types and amount of resources available. This process starts as a program is being developed, but continues through the evaluation activity. For example in nonprofit and grassroots organizations, the theoretical justification and backdrop for a proposed program initiative typically borrows eclectically from the diversity of theoretical models discussed in relevant literature, or by carefully articulating a chain of practical linkages between the chosen goal and the steps needed to accomplish it. In this way the theory can be either normative or causative, or in many cases, involve elements of both. But in each case it excerpts a parcel of the problem context into a manageable and delineated piece that can justify a request for resources that still makes sense in itself. The consequence of this process is that other appropriate theories or justifications for a program approach end up being set aside. Shaw and Crompton (2003) summarize the implications:

…the application of theory in evaluations without acknowledgement of alternative or wider theories is not only problematic, but may serve to decontextualize and depoliticize both services and evaluations. …choice of theory may be just as relevant and problematic to evaluation. In particular, failure to acknowledge wider theoretical or political agendas serves to remove such services and evaluations from political and critical analysis. As a consequence, arguments about the degree of freedom an individual should have over their own body is pushed aside in the quest for ‘better health’. ” (p. 201, italics in original)

As an evaluator I have often begun my work by developing and articulating program theory, although occasionally such theory is well articulated and understood before I arrive on the scene. In such cases I may take that theoretical backdrop as a starting point. However, sometimes my work involves exploring, identifying and introducing alternative theoretical frames in order to challenge the existing theory or its application, address new or emerging political contexts, or address important and potentially neglected aspects of context relevant to a reconceptualized understanding of the program. The next vignette explores an evaluation in which the existing program theory has been accepted and found to be appropriate by the
agency, its funders and most stakeholders. Yet the evaluation raised issues requiring a re-
examination of assumptions about the program and its theoretical underpinning.

**VIGNETTE 6.1.2 (NONPROFIT) – REFRAMING PROGRAM THEORY**

The community nonprofit organization has been delivering its well-known educational program in high schools across the region for over a dozen years. Although it has been very well received within the community, the agency has engaged me to evaluate the program with the view to updating its curriculum and overall approach. In addition to a new survey of students taking part in the workshops and a large database developed from follow up of hundreds of students who have completed the workshops in recent years, I conduct focus group sessions with a variety of community stakeholders, including teachers, counsellors, volunteers delivering the workshops, parents, and agency staff. My mandate has been defined as tweaking the program to bring it up-to-date, and making suggestions for how the content and format can be improved based on the perceptions of stakeholders.

My initial efforts point to some significant safety concerns for students taking the educational program. Several focus groups have suggested that the curriculum needs a complete overhaul, with basic assumptions challenged and reviewed. The survey data appear to show marginal gains in the key knowledge and attitude measures relevant to the stated program goals, and it is clear from the focus group with volunteer workshop deliverers that once they have mastered the core content, they adapt it in key and consistent ways. The implementation does not typically follow the original program design, and that is in part because that portion of the implementation is beyond the purview of the individuals conducting the workshops – the schools are not consistently carrying out their part of the
process. It is not clear that the schools are aware of this departure, or whose responsibility it is to arrange that everyone knows their part.

Based on some suggestions by participants in early focus groups, I bring together an external ‘expert panel’ to examine the curriculum in some detail, and to share their expertise concerning the current ‘state of the art’ as discussed in the literature. I facilitate this discussion over two meetings in which we review the curriculum page by page, and the panel contrasts the curriculum as it exists with ideas suggested by the current standards described in the recent literature. The expert panel is unanimous in suggesting a change in approach. The recommended changes extend beyond the limited mandate for my original study, but are consistent with the stated aims of the agency and broader network of stakeholders.

The report I prepare presents key findings from all of the different information sources, and encompasses both my original mandate – suggestions for updating the curriculum – and recommendations that describe an alternative scenario involving fundamentally changing the orientation of the curriculum. The latter has been developed based primarily on the suggestions of the expert panel, but also encompasses a detailed review of the relevant literature, providing the rationale for a new theory of change, and revised goals and processes for the curriculum. The suggested changes go far beyond ‘tweaking’ the curriculum, but remain consistent with the original goals and mandate of the program and agency. The recommendation also represents a significant expansion of the role of community partners in the workshops, entailing investment in a more inclusive approach to development, delivery and involvement in ensuring the safety of participating students.
In this vignette the broad program goals remain consistent, but the theoretical underpinning – its theory of change – is fundamentally altered. The chain of activities and the rationale for how they are done are transformed completely, although that result was not envisioned by my client, or part of my own intention in taking on the project. Indeed, the shift in evaluation emphasis was an emergent event, significant as a response to initial findings of the data collection, but also enabled by the client’s willingness to expand the scope of the evaluation to accommodate the political context and expressed needs of important agency stakeholders.

Reflecting on the theoretical basis of the program represented a significant challenge for me as an evaluator. As Bickman (1989) notes, the content knowledge expertise needed to do such work – to recognize the inadequacies of existing theory, to identify alternative theory relevant to the new circumstances and tailor it to the context – was significant. Despite having undertaken several other evaluations in the past in which I had examined and worked within this substantive area and with similar programs, building alternative theory suitable to the task stretched my capabilities and the limited resources set aside for undertaking what had been anticipated as a much more bounded and modest endeavour. What made it possible was using an expert panel who provided a nuanced and detailed critique of the program and its underlying rationale, identified cutting edge literature, and helped interpret findings from other information sources in a way that highlighted the deficiencies of existing theory, while suggesting where to look for a new framework.

What is particularly interesting about this case is that despite the existence of a clearly articulated program theory to guide the evaluation, the main outcome was the development of
new program theory, and a renewed vision of the program, its context, and the agency’s relationship with other stakeholders and the community. The evaluation impact has been far beyond what was anticipated by the agency, and this was specifically because I overreached my mandate to acknowledge contentious questions asked by committed stakeholders.

The evaluation effort in this vignette portrayed a process of reciprocal interaction between smooth theory and the ‘rough ground’ (using Schwandt’s phrase, 2003) of complex, contentious and contingent program practice. While acknowledging the value of theory in evaluation, it does not privilege theory over practice, and thus recognizes complexity but also ensures that theory remains a tool that is open to being amended on the basis of practice. In contrast, the use of logic models to represent theory seems to do the opposite. By establishing the program theory as a decontextualized and relatively closed model (in effect if not in design), it has the potential to exacerbate such unintended consequences as tunnel vision and ossification (Smith, 1995, pp. 284, 299). In many ways this is similar to the kinds of critique that have been made of performance measurement systems noted earlier (section 5.1 above) and in Funnell (2000). Schwandt (2003) describes this as:

… the hegemony of the functional in contemporary social practices – e.g. the obsession of education, management, healthcare, and the like with means at the expense of ends, with heightened interest in method and efficiency rather than purpose, and with questions of ‘how to’ rather that ‘why should’ (referencing Kincheloe & McLaren, 2000). (p. 361)

This recognition of the importance of context in the use of theory in evaluation is based on a vision of evaluation that is both interactive and dialogical (Van der Knapp, 2004; Schwandt, 2003). These ideas will be explored in more detail in section 6.4.

If theory is not just a guide to evaluation, but a product of it, and if we are to think not just of program theory but also of social science theory on a broader scale, then we need to reflect on how that interaction can take place. Another evaluation tool being developed and
refined by Schwandt and Dahler-Larsen (2006) and Abma (2007) is the idea of the ‘community of practice,’ a bounded community that has “its own norms, values, procedures, rules, ways of relating to others and so forth,” (Schwandt & Dahler-Larsen, 2006, p. 500), and in many ways can be considered to have its own ‘theory of practice’ that encompasses what to do “now, in this situation, given these circumstances, facing this particular person, at this time” (Schwandt, 2001a, p. 229). Vignette 6.1.3 explores some of the challenges of sharing a community of practice – a locally developed program theory and approach to designing and implementing a community-based program – with other communities.

**VIGNETTE 6.1.3 (GRASSROOTS) – SHARING A COMMUNITY OF PRACTICE**

The community’s innovative approach to working with individuals with a developmental disability has been figured out over an almost ten-year period. The agency developed its approach based on efforts of parents, and on the ideas of an academic who had given a talk at a conference. The researcher had discussed the importance of secondary disabilities – the idea that many of the key impacts for young people reflected not the initial disability, which could not be changed, but the effects of how others reacted to the symptoms of an invisible disability. Based on this idea, the local group had developed an approach that focused on inclusion – having the affected young people involved directly in conducting community education with service deliverers across the community. The intent was twofold – to address the skills and personal coping capabilities of the affected young people through intensive and supported practice over a period of time, and to give irrefutable examples to service providers of the benefits to their clients of using a different approach to interaction with them on a one-on-one basis. The central involvement by clients in conducting the
community education – often the very same clients that service providers had been dealing with on a professional basis – was a crucial element of the approach.

I have been involved in evaluating the program through its most recent funding cycle – an initiative funded over several years by both federal and provincial agencies, and building on the development work that had taken place over several years and prior iterations. By every measure I am able to examine and apply, the approach appears to represent an extremely successful effort on behalf of supporting individuals, and effecting profound and long-term change in participating organizations. Word of mouth – often by other community agencies that have participated in the workshops – has lead to significant demand from numerous communities to share the approach more broadly. On this basis, extension funding is obtained with the goal of expanding the model to six other communities; some similar, and some quite different.

My follow up evaluation of the efforts to expand to these communities shows consistent and comparable results in each location. Strong initial endorsement (each had heard of the model, and a local action committee had lobbied to be included in the expansion efforts), is followed by varying levels of success in establishing a local implementation, and within a year, each effort has become effectively ‘put on hold’ until the local community can regroup and try again. It is clear that in each location, members of the local action team have understood the model, appreciated how and why it works, and have been working to establish the conditions for their own implementation. The original community team remains available to mentor, discuss and support through several site visits and various means of support from a distance as local efforts become established.
A variety of barriers to local implementation are fairly easy to document – the lack of local infrastructure, reliance on volunteers who are working on the effort ‘off the sides of their desks,’ the challenge of building up a stable group of young people who can reliably undertake the activity of conducting the training once they understood it, the loss of momentum through a variety of delays in following up on initial training, and the loss of local champions and leaders to turnover within groups and changes to funding within organizations in each community. Even so, there also appears to be a more significant pattern to the challenge in sharing the model. The original community team had developed a sophisticated and powerful combination of workshop and training for those in other communities, and all those who participated in it directly appeared quite able to articulate the model, understand how it works, and appreciate what makes it powerful – the role of the young people with the disability who conduct the training. Yet when these individuals try to spread the model further within their own communities, they do not have either the practical experience or the credibility to pass it along, particularly because the model appears to contradict standard existing practices in how programming work should be done.

Other funders and administration within relevant service agencies appear to resist the substantial role of the young people in conducting the training. In two communities that do successfully establish local implementations for a short period of time, the key supporting staff (for very different reasons) are soon replaced by others who have not undertaken the original training. These individuals then proceed to try and take over conducting the community education and do it with a reduced or changed role for the young people involved. The established and understood model within the expansion communities has been that professional educators or clinical staff conduct community education, not clients. Unless
local individuals have seen the effectiveness of the model in action – either through the training conducted by the original community or by local individuals who have similarly undertaken the process – the model breaks down. People do not believe it. Funders will not give it credence. And those who understand and believe in the approach, still move on to other efforts when they run out of energy and enthusiasm trying to convince those who do not seem able to comprehend what is special about the approach.

**Reflection 6.1.3**

The evaluation portrayed in the vignette represents a dilemma in my practice. My efforts to understand the challenges to successfully transferring the knowledge and learning from one community to another stand as a testament to the difficulties in articulating “what works, for whom, in what circumstances?” (Pawson & Tilley, 1997, p. 85). The program is based on a clearly articulated theoretical model, has an extended history within a community, represents one of the most unambiguous examples of a locally successful community-based endeavour I have observed, and suggests an invaluable opportunity for many communities seeking ways to cope with a very challenging problem. And yet the model resists most efforts to transfer it to new settings. The situation also appears to represent one of those rare occasions in which an evaluator could provide a significant contribution by identifying what is missing, and answering the question, ‘how could other communities learn from and apply this model?’ Reflecting on the program as a ‘community of practice’ provides a useful frame for understanding some of these challenges.

Having been developed over an extended period of time, the local infrastructure – including the extensive level and variety of supports provided to the young people with the disability – clearly constitutes a significant barrier for other communities hoping to introduce
the approach from scratch. And yet, even if for only a short period of time, at least two other communities have been able to replicate enough of that infrastructure to establish a working version of the program that showed the potential of the model. My efforts to understand the barriers to establishing the program highlight the importance of the theoretical justification of the approach, and how challenging that is to pass along without the experiential component of seeing the program in action, and hearing directly from program participants.

In doing the training in any of the communities, the participants include a great deal of story telling: about their own backgrounds, the supports they need and receive, and what works for them. Part of this involves contrasting diverse stories from participants who experience similar situations quite differently. This helps get across the message of how individually tailored and emergent the supports are; that a one-size-fits-all approach is unsuitable for the range of problems experienced and the types of solutions needed. It also shows how the medium is the message – seeing the young people deliver the workshops and tell the stories makes it clear that many aspects of the program could not be delivered ‘to’ this group; that hiring participants to deliver their own program is the model. Such an approach contradicts our normative expectations of the relationship between programs and clients: between program delivery experts and professionals with sophisticated training, and program clients as helpless individuals who need things ‘done to and for them.’ The approach remains counter-intuitive, despite the evident successes, and does not fit easily within the frame of a funding and programming system that demands clear boundaries between programming activities and their consumers, and seeks to deny the community embeddedness of the activities, and the status of clients as citizens and stakeholders.
Beyond learning new ways to represent the program to decision-makers and other community stakeholders, my long-term interest is in finding ways to represent the program learning as extensions to and modifications of the theoretical framework from which it grew. The program theory is based upon a situated context of practice that is challenging to transfer to other communities even when there is strong impetus to do so. Part of the quandary is that the model itself meets resistance – it challenges and potentially threatens taken-for-granted assumptions and values about the world of program delivery (see Schwandt & Dahler-Larsen, 2006, pp. 501-502). I wonder how these assumptions can effectively be challenged unless we find an appropriate forum in which to articulate them, and possibly find other like examples of how people are challenging such assumptions with similar efforts? Being locked into instrumental questions about whether a theory or program ‘works’ does not easily afford opportunities to engage with such important questions, or allow the evaluator to take on the role of social critic, though this is an important task for which evaluators are often particularly well situated and potentially knowledgeable.

In conducting evaluations in small nonprofit and grassroots organizations, I have consistently come across opportunities to learn about programs and theories that reflect my exposure to the diversity of contexts in which programs are developed and delivered. Pawson and Tilley (2001, p. 324), Pawson (2003) and Lipsey (2000) emphasize the enormous potential of meta-evaluation to contribute to the development and examination of theory because of the variety of implementations across diverse contexts and themes. Such an opportunity certainly exists within the incredible range of programming undertaken by grassroots and nonprofit organizations. The context represents an enormous potential laboratory for studying ideas about programs and theoretical approaches to social problems,
and it remains an untapped reservoir despite the plethora of evaluation studies and reports generated and filed in the past two decades. While Schwandt (2000) emphasizes the caution needed when considering such extrapolation, I cannot help but see the potential of evaluation as a societal level activity, if we can move beyond doing evaluation in piecemeal bites – none of which satisfy, but all of which have something to contribute to a healthy society.

This section has examined some of the implications of the recent trend towards theory-based evaluation for examining programs and services in a context of devolution. Some of the key early critiques of theory-based evaluation included viewing it as challenging because it is time-consuming, expensive, and needs both extensive content knowledge (stakeholders) and theory construction expertise (evaluators). The nearly ubiquitous logic model seems to have addressed these concerns by putting the expertise into the hands of local stakeholders through universally available guides, although there have been recent concerns that such models over-simplify program theory and can not portray the complexity, multiplicity, or causal mechanisms of theories, or the contextual embeddedness of programs in the ‘rough ground’ of practice. Other concerns are that many programs – particularly those in small nonprofit or grassroots organizations – are unsuitable for the type of theory-testing envisioned for theory-based evaluation. They are not bounded or adequately stable, and indeed are emergent, changing, possibly volatile, and often do not yet have well-articulated or understood program theories. Further discussion in the section focused on evaluators’ roles in helping to construct, review and share program theories and ‘theories of practice.’

The next section continues this reflection by examining some of the ways evaluators are situated as participants within communities, programs and evaluations, and addresses the tensions inherent in juxtaposing stakeholder program knowledge and evaluator expertise.
Theme Seven: Inclusiveness

This section examines some of the implications and relevance of participatory evaluation approaches to the context of devolved programming. Over the past twenty-five years there has been a growing importance to acknowledging and accommodating stakeholders within the evaluation process. Diverse approaches to stakeholder participation in evaluation have been developed and promoted over this time, including stakeholder-based (Gold, 1983), fourth-generation (Guba & Lincoln, 1989), participatory (Cousins & Earl, 1992; Cousins & Whitmore, 1998), developmental (Patton, 1994), empowerment (Fetterman et al., 1996), utilization-focused (Patton, 1997), deliberative democratic (House & Howe, 2000), and responsive (Stake, 2004) approaches to evaluation. While each approach has slightly different characteristics and emphases, all share a view that it is both worthwhile and appropriate to facilitate enhanced participation in the evaluation process by those who are affected by a program. My focus in this section is on participation by stakeholders broadly considered – the various means by which stakeholders are included into any or all aspects of the evaluation process.

A key question asked by anyone exploring this area is “why do it?” Why include other stakeholders, when it is clear from the literature that doing so can be demanding of time and resources? Most look at this as a form of trade-off, in which the benefits of participation outweigh the sometimes considerable costs. Not surprisingly, the focus on evaluating devolved programs and services seems in some ways to turn this on its head.

The primary rationale for using participatory approaches to evaluation that has been discussed in the literature (Cousins & Whitmore, 1998, p. 6) is that they increase relevance and ownership, and thus promote evaluation use – they provide information that is locally
relevant and legitimate in a way that “constituencies are moved to act on what they know” (Freedman, 1998, p. 26). Ultimately, this ownership and enhanced use grow out of the perceived accuracy of the information and knowledge developed through the evaluation.

Following Garaway (1995), Cousins and Whitmore (1998) describe two broad approaches to participatory evaluation. Practical participatory evaluation (P-PE) focuses on practical rationales such as fostering evaluation use, while transformative participatory evaluation (T-PE) focuses on transformative rationales such as empowering community and in particular, those who are less powerful. Despite growing out of different traditions, the two approaches share significant attributes and reflect overlapping and complementary yet dissimilar purposes. P-PE is also capable of empowering people, and T-PE has potential problem-solving benefits. The approaches also share a prior epistemological assumption that the knowledgeability of local participants in programs is important, if not crucial for gaining valid understanding of program realities, and for making informed decisions. As will be seen in the discussions and vignettes below, the focus of most of my use of participatory practices when working with grassroots and small nonprofit organizations has been quite pragmatic. While at times the context enabled transformative or empowering emphases, the day-to-day use of participatory techniques reflected a practical approach encompassing instrumental, educative and symbolic (Cousins & Whitmore, 1998, p. 6) intended effects.

Cousins and Whitmore (1998, pp. 10-11) also propose three process dimensions for distinguishing important characteristics of collaborative inquiry. The dimensions – control of the evaluation process, stakeholder selection for participation, and depth of participation – provide a useful lens for reflecting on the nature of participation by various stakeholders in the devolved evaluation process. Figure 6.2.1 below replicates and adapts Figure 1.1 from
Cousins and Whitmore (1998, p. 11), with several key changes that will be discussed in more detail below. The adaptations are reflected in each of the three dimensions represented.

**Control of the evaluation process** ranges from *researcher controlled* to *citizen controlled*, the latter changed from the original ‘practitioner controlled.’ **Stakeholder selection process** has been renamed ‘who participates’ and ranges from *all legitimate groups* to *selected groups*, with the latter changed from the original ‘primary users.’ **Depth of participation** has been renamed ‘levels of participation,’ and ranges from *deep participation* to *token*, with the latter changed from the original ‘consultation.’ These adaptations were first developed and presented in Harper and Hinbest (2002).

**Figure 6.2.1: Process Dimensions in Collaborative Inquiry**

**Process Dimensions in Collaborative Inquiry**
(based on Cousins & Whitmore, 1998)
Vignette 6.2.1 explores several characteristics of the dimension ‘control of the research process.’ This encompasses the dialectic of control between evaluator, funder, practitioner and community, and as Gregory (2000) and Themessl-Huber and Grutsch (2003) note, how the control over the evaluation process can and does change over time.

**VIGNETTE 6.2.1 (NONPROFIT) – PARTICIPATION REVOKED**

My official client – a provincial government department – has approved funding to support the efforts of an evaluation advisory committee. The committee has representation from a variety of stakeholders, including nonprofit organizations delivering the program, administrators from the funding department, members of a provincial advocacy group interested in the programming area, and young people representing the program target group – program-eligible individuals. At the inaugural meeting of the group four weeks into the evaluation study period, I share my initial draft of data collection instruments for comment and discussion. Some members of the group express surprise that I have already developed draft instruments, as they had hoped to have a larger role in designing the study and setting its direction. I emphasize that the instruments are drafts, and can be changed, and indeed, I wanted to have something concrete to begin working with so that our time would be used effectively. I also note that the timeline for the study is relatively short, and that my contract with the funder calls for me to have the instruments field-ready within another three weeks.

As the discussion proceeds, members of the group get into the spirit of rolling up their sleeves and critiquing the draft instruments. As there are several groups from whom data will be collected, there are several instruments, and we break into smaller work groups to focus our efforts. After several hours, there appears to be substantial progress in working through some of the key challenges in refining our approach. It seems clear to me that the knowledge
of the program delivery context that most people have brought to the committee will prevent a variety of pitfalls, and the committee seems buoyant and in good spirits at the progress made. As we prepare to take a short break, young people from one of the groups with several ‘program eligible’ individuals raise a question. They want to know where we have addressed the issue of access to the program, as most of the data and indicators being developed focus on looking at impact of the program once people are in it; there seems to be a gap in that we are not examining access and the application and approval process.

I suggest that this could be an important issue, and look to the funder’s representative for comment. Her response is clear, that access is not on the list of approved evaluation questions developed from the department’s logic model, and that it is too late to add such an issue now. I wonder whether she is worried that if we add this issue I might expect either more time or more resources to conduct my part of the study – that I would try to lever more money for the contract, perhaps – and so I comment that given the current design, it would not be difficult to include the issue of access for some of the data collection, although to address it fully would mean adding a new information source and possibly expanding one survey group. I ask the young people if they have any specific reasons for identifying this issue, and they and the advocacy organization representative share several anecdotes that paint a picture of how organizations only accept clients who are most likely to succeed – that the people most in need are not getting into the program because agencies seem reluctant to take them on.

The funder’s representative still appears to be disinclined to add the issue to the study, but in the face of the growing enthusiasm and endorsement of most committee members, agrees to adding a few questions to examine the issue, but she is not willing to
extend the survey design. The meeting concludes on a positive note, with a plan to have the revised instruments and next steps shared with the advisory committee for further feedback shortly, and sign off by the target deadline set by the funder.

Over the next two weeks redrafted instruments and evaluation plan are shared with the committee, and useful feedback provided. As we move into the next phase of data collection, contact with the committee is infrequent – several e-mails are sent out with updates, but no meetings are arranged. As we move toward the end of the first data collection phase I ask the funder’s representative when we can schedule our next meeting with the advisory committee. I am told that the funding to support bringing the committee together has been withdrawn, and that the group has been ‘disbanded.’ She explains that we really didn’t need their input any more anyway, so I just need to write the report for review by her and the program management committee, and not the advisory committee.

As I conduct interviews, analyze the data, and write the report, it becomes clear that the access issue is a significant one for the program. It has been named as one of the three top concerns by each of the respondent groups. My first draft of the report to the client includes a section specifically on the issue, laying out the concerns, and some possible implications and strategies. My client calls and tells me that I need to remove this report section, as it is not in the contract, and not listed in the issues and questions outlined in the original Request for Proposals (RFP). I refer to the advisory committee’s desire to have this issue included in the study, and am reminded that the committee no longer exists. I am able to negotiate a way of including the issue in my final report, although with less prominence and a substantially reduced emphasis.
The vignette portrays a situation in which the impetus for a collaborative process appears to have come from the funder, although it may have reflected lobbying by advocacy groups prior to the start of the project. From my perspective as the external evaluator, I was not anticipating the existence of this advisory committee, and was informed about its formation only a week prior to the first meeting. I have no reason to believe that the funder’s representatives were not sincere about their desire to elicit participatory input for the study design and implementation, and indeed another facet of participation was the known contract requirement that the study employ target-eligible youth for data collection activities. The source of the withdrawal of the committee’s participatory voice remains unclear, although the subtext appears to point to concern with how additional evaluation issues were raised by the committee. In effect, a tentative approach to power sharing was attempted, and quickly retracted when the advisory committee tried using its influence to alter the study design.

The context represents an early phase of the devolution process, in which the delivery was conducted by local nonprofit organizations, but the evaluation was still mandated and implemented centrally. As such, the report was subject to public release, and elicited a high level of managerialism in the form of significant redrafting of how issues, wording and results were portrayed. The continued existence of the advisory committee would have facilitated my ongoing negotiation with the funder, and I tried to draw on that resource, but to no avail. As a form of participation, the situation served to show me some of the potential benefits of making the evaluation process more public and participatory.

Cousins (1996) suggests that the exclusive reliance on external contractors for evaluation appears to be declining in the context of austerity and increased demands.
regarding the accountability and responsiveness of applied research. Such a decline may also reflect the context of devolution and efforts to control the transparency of evaluative outcomes. Either way, from my practice it is clear that the role of external consultants is changing, particularly with regard to collaborative evaluation approaches. Elsewhere, Cousins et al. (1996) report on a survey concerning collaborative evaluation and note that what many evaluators refer to as undertaking collaborative evaluation seems to resemble conventional stakeholder-based evaluation approaches, and not necessarily participatory evaluation. Situations in my practice related to the vignette suggest to me that one way of analyzing such a finding is to note that the Cousins and Whitmore (1998) seem to conflate the role of evaluator and funder. The vignette portrays a situation in which the evaluator role is relatively unambiguous, and yet even in a minimally participatory setting, the unequal and yet shifting power of the evaluator relative to the funder seems transparent after the fact. At the time, however, I would suggest that from the perspective of the nonprofit organizations and advocacy group, it would have been hard to distinguish the evaluator from the funder. As I had no means or mandate for following up with those groups, they remained unable to discover the impact of their efforts other than by reading the final report, which eventually became a public document. Given the significantly more ambiguous role of the evaluator whose direct client is the nonprofit organization, intervening federation or agency, or some coalition of groups as has been described in previous vignettes, the potential complexities of establishing and maintaining participatory partnerships appear to be enormous.

When Themessl-Huber and Grutsch (2003) discuss the shifting roles and hats worn by stakeholders in a participatory evaluation, they argue that the locus of control is shifting to participants and away from evaluators, and implicitly this seems to equate the evaluator’s
role as equal to or at least identified with the funder’s role. While they usefully note that participatory stakeholders are not powerless, and indeed, express their power overtly through control, and more obliquely through resistance, they still seem to underestimate the incongruities of multiple players and interweaving roles and mandates.

Looking back to Figure 6.2.1, the vertical dimension ‘control of the evaluation process’ originally ranged from researcher controlled to practitioner controlled. The adapted label for the latter is citizen controlled, reflecting the way that those involved in making decisions within a participatory context include not just practitioners, but other stakeholders as well. Although not reflected in changes to the dimension in the figure, the label researcher controlled also seems to reinforce the conflation of the researcher with the funder. Perhaps another dimension would help to reflect the variability in the independence of the evaluator with respect to the funder. On the other hand, in many, if not most contexts of devolved evaluation, there may be no intervening ‘evaluator’ between the funder and the local practitioner overseeing the local end of the evaluation efforts.

A more general issue raised by this vignette concerns how we determine whether an evaluation is participatory. The literature appears to treat participation as somewhat of an all-or-nothing proposition. Harper and Hinbest (2002) argue that as a context-specific process of engagement, participatory evaluation must often deal with obstacles to practice and contradictions between democratic and methodological goals of participation and ideals of efficiency, as well as competing demands between implementation and evaluation. So the various elements of participation may be quite diverse in practice, even within any one evaluation project. When we consider that it can encompass study design, framing evaluation questions, designing and conducting data collection, analysis of data, development of
reporting processes and implementation of findings, this potentially represents an enormous commitment of time, resources and expertise on the part of practice partners. Indeed, many of the initial efforts to introduce participatory elements to evaluation, particularly in development contexts, focused on employing local knowledge experts to either collect data or assist in identifying important meanings and interpretations of data and issues (Freedman, 1998; Brisolara, 1998; Cousins & Whitmore, 1998). In my experience, it is seldom that more than a few participatory elements are possible within any one project. Cousins (1996) addresses this issue by positing a scale of forms of participation, ranging from full to silent to being a general advisor, and acknowledges a frequent mismatch between the estimates for workloads and priorities between evaluators and participating staff.

Finally, Cousins and Earl (1995) define participatory evaluation as “applied social research that involves trained evaluation personnel (or research specialists) and practice-based decision makers working in partnership” (p. 8, italics in original). This definition appears to limit how we view the stakeholders within the practice context. In particular, it is challenging to identify whom in the practice context is a ‘decision maker’ – and given the devolution setting, the lines between administrator, front line worker, volunteer and even client, can blur. The next vignette explores some aspects of how the nature of participation in an intentionally inclusive setting can encompass more than a one-dimensional transfer of either expertise or local context knowledgeability.

**VIGNETTE 6.2.2 (GRASSROOTS) – EVALUATOR OR PARTICIPANT … OR BOTH?**

In supporting the evaluation efforts for a small one-year innovative program, I am faced with the need to talk to and learn from a group of participants who have developmental disabilities. Although they are for the most part relatively high functioning, some of the
participants in the program have limited literacy skills. Program coordinators have suggested that if I simply come to a group session and try to conduct a focus group, I will have little success – one or two individuals might talk to me, but most would remain silent. The coordinators do not feel comfortable in taking on the role of conducting any interviews with the group, but instead, suggest that I simply come to the group sessions for a while, build up a relationship with participants, and let them become comfortable with my presence.

The agency’s programming focuses on establishing inclusive approaches to working. Attending information sessions and presentations I have been impressed with the way that participants are never referred to as ‘clients,’ and indeed, at one community presentation conducted by one participant and one staff person working together, the first question from the audience was to ask which individual, if either, was participant, and which was staff or volunteer. My first encounter with the group session reinforces this impression, as I notice that every staff member of the agency makes an appearance over the course of the two-hour gathering. People come and go, acknowledge the group and are acknowledged in return, move on to other meetings or phone calls, give regrets or plan to attend some event in the coming weeks. People ask how things are going – both staff and participants – and seem to know a great deal about everyone else’s family situation, plans, outside activities, and lives.

The meeting itself confirms this apparent lack of boundaries between staff, volunteers and clients. To be sure, at least one staff member is a former client, most were volunteers prior to being hired, and all of the current volunteers are former or concurrent clients of this or another agency program. Because of my work with several other programs at the agency, I know which individuals are currently on staff, but if I had not known, I would not have been able to figure it out from the discussion and interaction. The participants joke and talk and
share what they are coping with, and so do the staff and volunteers. When it is my turn to speak I am not introduced by the coordinator, but simply say that I am a researcher with the agency, and I’m interested in understanding what they do and how this project works.

As I attend weekly meetings of the group for the next six weeks, I am drawn into the lives of participants and staff, share stories of my own foibles as parent and family member, and get to know people and let them get to know me. At various points as we go to outings or for a walk to the park to vary the scenery, most individuals take a moment to speak with me individually, asking questions about research and evaluation, but also just checking in to see if I’m all right; if attending the group is helpful for me, and how my life is going. I am touched by the genuine caring of the people I am meeting, and learn to care about what is going on in their lives too.

By the end of the six weeks I have raised the issue of getting feedback about the program from the individuals attending the group. Although I feel that a more traditional ‘focus group’ might be possible at this point, I am interested in trying something different. The group brainstorms with me about possible questions to ask, based on samples that I provide. We come up with ten questions we’d like to know about, which we write at the top of ten flip-chart sheets that are taped to the walls of the meeting room. It seems that everyone in the agency descends on the meeting, and we all, including myself, spend the next hour and a half walking from sheet to sheet, talking about the questions, reading other peoples’ responses, and helping one another think about, articulate and write replies to each question. Everyone has their own differently coloured felt pen to write with, and I have a sign up sheet, but most people sign their own comments anyway so I know who has said what.
The answers given to questions are consistently respectful of others, but include critical comments as well as positive ones. In talking about my goals over the six weeks, I have emphasized that the goal of my research is partly to help the agency plan things and do them better, but also partly to be able to share this program and experience with people in other communities who might be interested in developing a similar program. Both staff and participants seem excited about the possibility, and look forward to finding out how they can contribute to others. I am not surprised when some of the key benefits of the program appear to be that it develops in participants a belief that their input is worthwhile, and further, that they are capable of helping others, and in so doing, help themselves. Participating in the evaluation also seems to have helped participants learn that they can do research, and perhaps help others they have not yet met.

**Reflection 6.2.2**

This vignette portrays a setting in which a participatory approach was used for virtually all aspects of agency programming. As such, a condition of my being hired to assist with evaluation work was that I use inclusive methods and decision-making in all ways that made sense and were feasible. My planning was participatory within the agency, but I had little direct planning contact with either the funder or other community partners. The short one-year duration of the project precluded more intensive participatory planning efforts, although such activities did exist for other multi-year programming within the organization.

Program participants seemed to have little awareness or concern that this was a ‘program’ – they came to the organization one or more times a week, but when I asked them about the ‘X’ program, a consistent response was “What’s that? – It sounds interesting.” My efforts to ask them questions about the program did more to bring the reality of the usually
unstructured and sometimes-educational activities they were undertaking together and bound them in a way that they might consider it a ‘program.’ Much as Kushner (2000, p. 10) observed, it seemed to me that for many of these participants the very concept of a ‘program’ was little understood and for the most part irrelevant to their ongoing lives. And yet their participation in the program and even the evaluation activities had meaning for them.

My own efforts to understand what the program meant to participants likely had as significant an impact on my life as on theirs. The activities of the mutually given and taken ‘focus’ group seemed to be a delightful spark to staff and volunteers as well, as they continued to reflect on the discussion afterwards and in other contexts, many of which I have heard described but have not examined systematically. I am convinced that if I had not participated in the discussions and group activities as I did, I would not have been able to obtain the kinds of information I obtained and was able to portray in the evaluation report distributed within the agency and sent to the funder.

Mathison (2001) describes a case study in which the participatory evaluator is a “genuine” stakeholder. She uses examples to talk about how as a genuine stakeholder and an evaluator in the same context she was required to choose a dominant role in any given point in time. This entailed feelings of frustration and conflict, and was a constant source of reflection for her. It involved learning to reflect on the most appropriate timing of making contributions, and using her own emotional responses as data to help understand and identify potential sources of information relevant to the study. The kinds of intensive personal interactions I have experienced in the process of blurring the evaluator / participant roles have certainly introduced for me a clearer understanding of my own responses and reactions as relevant to the evaluation process, and that even when they are not ‘data’ for the study,
they represent important means by which I can understand and build meaning with the other participants in these processes.

Reflecting on the dimension of Figure 6.1.1 labelled ‘who participates,’ the program’s staff, volunteers and clients were all involved in participatory aspects of the evaluative process. Although I wrote the report based on a variety of sources in addition to the one described in the vignette, it was developed through planning with staff and volunteers, the data collection involved diverse local stakeholders, and the follow up involved many as well. The time was short for focusing on a great deal of education in the evaluation process, although our focus group discussion did involve talking about the implications and challenges of trying to develop similar programming in other communities, and how this program might become sustainable locally over a longer period of time. This awareness by clients of the larger context of programming has carried through to their participation in other agency programs, and in particular to their potential movement into volunteering roles with the agency. In this way it seems important to look beyond the short-term limitations of individual projects – a particularly challenging aspect of grassroots delivery – and think about participation as an ongoing series of efforts that will span multiple projects.

The next vignette explores another facet of the process of involving stakeholders in participatory evaluation efforts within a devolved programming setting – coping with the challenges of turnover and inconsistency among the constellation of possible participants.

**VIGNETTE 6.2.3 (NONPROFIT) – REVOLVING MEMBERSHIP**

Evaluating a community level program that involves several agencies working together as a coalition, I arrive at the monthly meeting of the coordinating group seeking to finalize plans for data collection that will involve representatives from at least four
organizations in collecting data within their organization and one or more others. In advance of the meeting I circulate the final drafts of data collection protocols and instruments revised at the last meeting. I know that people are busy, but I am hoping that they will have reviewed them prior to the meeting so that we can quickly sign them off and move on to some of the many next steps on our agenda.

As people arrive and the meeting officially begins, I see three new faces among the seven around the table. As we introduce ourselves, the three state that they are filling in for or replacing others who could not make it to the meeting. Each of the new members seems to know several others attending the meeting, although no one knows everyone. Two of the three new people and one who had attended the last meeting have not yet seen the ‘finalized’ materials, so we begin going through them quickly. By the time we answer questions and resolve a few minor changes, we have used almost half of our allotted time for the meeting.

I find myself being frustrated by the slow pace, but also aware that I am and should be grateful for the full turnout to the meeting. I am not clear about whether the new agency representatives plan to attend regularly, whether they have replaced the individuals who could not attend this evening, or if I may never see them again. As we take a brief break for refreshments, I chat with the new members and ask about the people they are replacing, and when we might see them again. One states that she is only here for the one meeting, unless the usual representative is unable to attend in the future, and then she would again try to fill in. Another is clear that he has replaced the old representative and plans to attend for the balance of the project. The third surprises me with her answer – she is here for a while because it is ‘her turn.’ Others in her agency will get their chances as opportunity allows.
In her article “Problematizing Participation,” Gregory (2000) deconstructs the concept of participation and how evaluators have used it to describe their work. She challenges the idea that participation is necessarily good, she examines the impact of power imbalances on participation, and she reflects on people's motivations to participate, motivations to invite the participation of others, and the possibility of being co-opted as part of a participatory process. Gregory notes that the processes for selecting those who will participate in evaluations and the modes of representation are necessarily political acts that can involve gatekeeping, intensive levels of cultural knowledge, and knowledge of the range of stakeholder groups. She also notes that all of this is an emergent process, and the range of stakeholders may change over the course of an evaluation.

This exploration of the problematic nature of representation and participation seems particularly relevant to the devolution context. I have found that there is usually a trade off in identifying and negotiating the right level of representation both within an organization and especially among them when building an interagency committee or coalition. It can be extremely important to establish in participants a comparable level of authority to represent the organization and to speak on its behalf. Yet for most issues the representative must also be close enough to the front line to know and accurately represent implementation issues. Grassroots organizations often rely on volunteers for this, and they seldom have protocols on how to represent the agency, or ensure follow-up among staff or other volunteers. Further, as anyone who has tried to establish a long term community representation for a group or committee will appreciate, executive directors are usually not interested in attending meetings with front line workers or volunteers from other agencies. The status inconsistency
can make coalitions fragile. If there is too much delegation of representation for individual
meetings, the nature of the coalition changes, as all are soon delegating, and the power of the
group to effect change may be reduced or compromised. The balance between selection of
participants and self-selection is an inherent aspect of the power trade off in participatory
evaluation, and even if for practical or emancipatory reasons self-selection may be
encouraged, its implications for conducting the evaluation need to be acknowledged.

This self-selection has been a major challenge and strategy in my experience with
encouraging participation in evaluation among grassroots and nonprofit organizations.
Mathison (Ryan et al., 1998, p. 115) suggests “the prior question is... who are the stakeholder
groups? We sometimes assume the identification of such groups is fairly easy, when in fact,
it may not be.” And in my experience of being the consultant going into a community from
the outside, I must rely on communities to help in the process of identifying stakeholders and
building enthusiasm for participation, particularly when peoples’ plates are already full. One
challenge in this political process is not being aware of the various power balances within a
community; by selecting one individual, I may guarantee the nonparticipation of several
other key community members. Selecting participants can be a very political act.
Encouraging self-selection does not depoliticise this process, but it can defuse some of the
local politics, make them more transparent to participants as they justify their own choices,
and leave the political side of the process to insiders. On the other hand, part of my own
focus must be on working to identify those who are not represented – particularly those
without power to demand participation – and to advocate for their inclusion or actively try to
represent their interests or perspectives (Ryan & DeStefano, 2001; Greene, 2001). Referring
back once again to Figure 6.2.1, the dimension ‘who participates’ was renamed (from
invited or self-selected to become involved, and supported in their participation.

One of Gregory’s (2000) concerns with Patton’s (1997) utilization-focused evaluation is that it defines who should participate in terms of whether or not they are end users of the evaluation. She finds this to be problematic in the sense that some people have a stake in the program or its outcomes or the process of conducting the evaluation and yet might never be an end user of an evaluation product. Beyond the challenge of identifying who might be an appropriate end-user for evaluation work, it is very difficult to anticipate all of the many ways that evaluation findings or reports might end up being used. I have tried to promote their use by those who are designing programs – whether or not they are directly applicable. In the previous vignette, the evaluation process ended up being an impetus for agency-wide discussion by raising questions and issues, but I doubt that the evaluation report itself had much direct impact on the lives of the program participants. Its impact on the agency as a whole provided less direct impacts, and may indirectly have affected the work of community partners.

Once we have identified the stakeholder groups, and determined whether they are chosen by the participatory facilitator or self-selected – assuming that they remain constant over the course of the study – we have to think about the form of representation. In my practice this becomes an issue not just because there is turnover among those who participate on steering committees or evaluation workgroups. Often, participation in such groups is considered to be a perk or a bonus to people within the organization, and not something that is given to only one person as a specific task to deal with. So in an effort to be ‘fair,’ organizations sometimes rotate participation at committee meetings among two or more staff...
or volunteers. This may contribute towards good morale among the organization’s staff, but it can represent a challenge to ongoing continuity between meetings. Another rationale for this kind of approach reflects the organizational reality of grassroots and small nonprofit organizations. The flattened organizational structure of such agencies means that there are few opportunities for front-line staff to gain experience that can prepare them to take on management and administrative roles within organizations. Participation on such committees can be seen as an opportunity to build important experience by junior or front-line staff or volunteers. As a training ground it may be effective, but it also represents a way of passing along that training function to other community members, and perhaps to evaluation facilitators as well.

This educative aspect of working within participatory contexts is a recurrent theme from my practice, and leads to a focus on how the capacity to participate may be taken for granted, when it may be a new experience for most of those involved. The next vignette explores a consistent experience from my practice – the trade off between supporting effective work, and having the patience to let people build, demonstrate and articulate their own knowledge and knowledgeability. Allowing people the opportunities to make their own mistakes can challenge efficient use of scarce evaluation resources, but with a longer-term focus we can hope to enhance capacity for future participation.

**VIGNETTE 6.2.4 (NONPROFIT) – BUILDING LONG-TERM CAPACITY**

I hear great pride in the voices of my two local evaluation colleagues in a Northern community that I am supporting from a distance. They have attended two workshops in Vancouver, and have enthusiastically embraced their roles as local evaluators for the eight-community demonstration project. After working with me and representatives of two other
Jerry: The survey really looks great. I see that you have sent me a copy of it with some numbers written in after each question. Can you explain to me what they mean?

Alice: Well, those are the total number of people answering that question.

Jerry: So for question six where we had the scale from very useful to not at all useful…?

Alice: Ten said very useful, fifteen said useful, eight said somewhat useful and three said not at all useful.

Jerry: So you had thirty-six responses to the questionnaire?

Alice: Thirty-seven. I think one person didn’t answer that question – you called that a non-response, right?

Jerry: Yes, that’s what we call it. So that’s good – I see how you have summarized the response frequencies, and that’s a very useful start. I don’t see the rest of the completed questionnaires in the package you sent; did you send them separately?

Terry: No, they were destroyed, like you told us.

Jerry: Pardon me? You say I asked you to destroy the questionnaires?

Terry: We listened very carefully when you were talking about ethics and confidentiality and making sure that we respected everyone’s privacy.

Alice: Yes! We told everyone that the survey was anonymous and confidential, and that we would destroy the surveys after we counted them up. So after we finished summarizing the data last week we burned them.

Terry: It was fun – almost like a ceremony! It was so cool.

Jerry: So, ah, do you have another sheet where you transferred the results from each questionnaire to a record – to an individual summary of the responses for each person who was surveyed? You did summarize those, didn’t you?

Alice: We sent you the summary – you’ve got it. Did we do something wrong?

Jerry: Usually we don’t destroy questionnaires until we are sure we have obtained all of the information we need from them. We can remove any identifying marks to make sure
they can not be linked to the person who filled them out, but we wait until we can be sure all the data are accurate, and entered into a database of some kind, and checked again. And it’s important to capture how the different items relate to one another, so we can find out if males had different responses than females, for example.

Alice & Terry: Oh!

Terry: Maybe we could get them to fill out the questionnaire again?

Jerry: I think we can make do with this summary – this is just another of our lessons learned, eh? And your intentions were great – the people in your community should be very pleased with your respect for their confidentiality.

Alice: We didn’t know…

Jerry: And that’s my mistake. I didn’t give you enough detail about what to do with the surveys. I think you two are too efficient – I wasn’t expecting you to be done this soon, and I didn’t give you the information you needed quickly enough. But next time we’ll all know better, right?

Alice & Terry: You bet! So what’s next?

After the call, I shake my head in disbelief. I am surprised that Alice and Terry did not understand such basic elements of research, but realize that I have not covered this in enough detail in the workshops – I had assumed that those technical details would be more within my domain. I doubt that others will have destroyed questionnaires, but proceed to make individual calls to other local coordinators to establish data management protocols.

**Reflection 6.2.4**

One of Gregory’s (2000) concerns was that much of the work by those writing about participatory evaluation underestimates how difficult and challenging it can be to learn how to be an effective participant in a participatory project. Vignette 6.2.4 describes a situation in which participants have successfully grappled with many facets of participation in the effort: coordinating among diverse community interests, respectfully representing the interests and needs of their community members and project participants, enthusiastically engaging in the evaluation endeavour with the view to understanding and improving the project as well as
contributing to knowledge as part of an inter-community coalition. The participants’ stumble reflected the challenge of identifying the boundaries of diverse roles – the technical from the representative; planning and designing the participatory efforts compared to the implementation of the activities. My challenge was in respecting the enormous progress and commitment represented by the activities of my colleagues, while turning it into a learning situation for myself and for their future participation. I needed to maintain their enthusiasm while encouraging them to treat my input as a consultative one – when they are in doubt, perhaps I could help and they could call me, particularly with technical questions.

The vignette does not show the progress over previous iterations of a similar project, in which the local independence of the on-site team included designing and implementing such surveys without coordinating them at all with other communities or other off-site team members, myself included. The progress represents learning by both the local community participants and myself as the facilitator of the participatory projects.

Examining the changing nature of power between evaluators and local communities, Wallerstein (1999) discusses how the relationship has been changing over the past two decades to a model of evaluation ‘with’ communities as opposed to ‘on’ them. She notes that this involves contradictory and sometimes problematic relationships in terms of their differing levels of power, including evaluations that are intended to be participatory in nature. Wallerstein's concern is that when evaluation is mandated and imposed by funders, evaluators have to be aware of the implications this has for their relationships with local community members, and their attempts to involve them in participatory processes. She notes that there is a tendency to conflate participatory processes with ideas of empowerment, and
that this may be problematic unless we learn how to ‘walk the talk,’ recognize our
differential power in the situations and actively work to address them.

In the vignette, the subtext to my work with the local communities was that despite
their buy-in and enthusiasm, as well as the non-accountability focus of the evaluation efforts,
the participation in the evaluation was mandated by and a condition of participation in the
project funding. In some ways my support for the local project activities was welcomed and
taken in good faith. In others, it was perceived by local participants as either unwanted
imposition of external control on their activities, or an equally unwanted devolution of
evaluation responsibilities that required local participants to learn lots of new skills and do
work for which they perceived themselves to be untrained and under-compensated. Some
asked me directly to “do the evaluation part” – again, something that because of distance and
cost, I could not have taken on if I had wanted to. So from the perspective of some people in
these participatory efforts, the evaluation was not just done ‘to’ or ‘on’ them, but they were
expected to do it to themselves – another example of imposed self-regulation.

The inherent contradiction in the way that the requirement to undertake evaluation is
often imposed even though it may employ a participatory approach in practice represents a
significant challenge for the evaluation facilitator. My approach to reconciling these
seemingly conflicting ambitions is to emphasize two transformative facets of participatory
evaluation (Cousins & Whitmore, 1998, pp. 7-9). They are transformative in that they
recognize and in some ways address differential power and inequality within the contexts.
One is in recognizing the unique knowledge creation capabilities of those within local
communities – their knowledgeability makes contributions to evaluation efforts that only
they as local participants can make. Since typical evaluation work imposes a knowledge
framework and indicators upon local settings, I emphasize that a participatory approach represents an opportunity to make sure the information collected and standards of criteria for measuring progress are locally relevant and appropriate. I also emphasize that even with the best of situations, it is often challenging to translate local knowledge into forms that centralized areas and funders can understand and are either able or willing to appreciate (Yanow, 2004), and so it is very practical and transformative for community members to take advantage of those rare opportunities available to them.

The second transformative emphasis I make is working to connect the activities and the evaluation in the local community to what is happening in other communities. In the vignette, this was made easier because the project itself provided direct linkages to other projects and other communities that were undertaking complementary activities. In one sense I am helping to provide a clarified vision of the broader implications of what they are trying to accomplish in the community. In another, I am drawing on the local initiative and commitment on which the project depends (Sergiovanni, 1992), and linking it to a broader moral imperative that makes what they do part of a larger initiative. This transformative emphasis reframes the project and takes local disparities and turns them into strengths in how communities can contribute to other settings and problem contexts. It also provides a practical benefit of broadening the stake upon which local communities are stakeholders, and provides a means of making the learning that emerges from the project more meaningful.

Examining these issues of power and local community participation in this way has led me to re-evaluate my understanding of the empowerment evaluation of Fetterman et al. (1996). Patton (1997b, 1997c) notes that empowerment evaluation concepts and techniques overlap significantly with such ideas as participation and collaboration. Empowerment
evaluation has become an important addition to the complement of participatory approaches available, but as Patton (1997b) notes:

One doesn’t get very far into the political culture of empowerment advocacy before encountering the question of whether one person can empower another, for to so claim is elitist and disempowering of the self-empowering capacity of each individual. The philosophy of empowerment evaluation is that no one can empower another; people empower themselves. But, in practice, the language can be tricky. (p. 158)

My reading of empowerment evaluation is that the approach so strongly emphasizes local autonomy and control of the evaluation process that it can end up removing the shared vision by means of which the results and outcomes of such evaluation could be extended to help and inform work in other communities. What appears at first glance to be enabling communities to set their own appropriate agendas for evaluation, actually can amount to abandoning them to their own devices. In other words there is a role for the facilitator of a local evaluation that encompasses building that vision, raising the bar, and clarifying for local communities how what they are doing relates to and supports what is done elsewhere. In the local communities and nonprofit organizations I support, I provide some technical support about evaluation approaches and methodology, but often the key aspects of my involvement are my work in connecting communities, facilitating discussion among them, and asking questions that will help them put their own work into the context of what is done and needed elsewhere. This helps them to better understand what it is they are doing, but it also helps to develop the community level equivalent of Sergiovanni’s (1992) moral authority – helping people to understand why they are doing what they do. In many ways they already know, but they need to be reminded that their commitment is not just to an individual endeavour, but also to a broader project, community, group of communities and society.

Applying this same transformative perspective to my own work as an evaluator allows me to see beyond the individual project even when it is not directly connected to
others. Capacity building and education about evaluation are long term processes for which the payoff may be years down the road, and may be beneficial to another evaluator and not to me. I may never learn of the impacts of my capacity-building efforts, but I will see it when I come across cases in which the preparatory work of other participatory evaluation facilitators has made my job easier and more effective.

Whitmore (1998b) describes participatory evaluation as a “complex, time-consuming, and risky business” (p. 97). She notes what she terms five ‘essential ingredients’ to making participatory evaluation work: 1) having a receptive context (with everyone on board and an open and democratic organizational climate), 2) commitment to participation and faith in the capacity of participants to meaningfully contribute, 3) adequate time, as the process can’t be rushed, 4) facilitation and interpersonal skills to work collaboratively, and 5) engaging stakeholders in the participatory process and in capacity building (pp. 97-98). My work with nonprofit organizations seems to belie Whitmore’s essential ingredients in many ways. While I have occasionally found a receptive context, I have usually had to start from a position of scepticism with the process, and cope with a milieu in which no matter how participatory, the process that enabled it also made evaluation obligatory. I have rarely found universal commitment to and faith in the participatory approach in the projects, but rather developed that as part of the participatory process itself. There has never, ever, been enough time to give the participatory course its due. My ability to read and understand the complex internal and inter-organizational contexts, and to facilitate successful participation has repeatedly been faced with unanticipated challenges and my own limitations, to which my recourse has been admitting my mistakes and my shortcomings, and then working to build faith that I too am learning to facilitate, participate and build capacity. And engaging stakeholders and
building capacity typically ends up being as much outcome as condition; never completely achieved, but for the most part reflected in a two steps forward, one step backward dance towards the goal of a full, successful participatory endeavour.

My point here is that despite all of the challenges of conducting participatory evaluation in grassroots and nonprofit organizations in a context of devolution, and contexts that never have met Whitmore’s five essential ingredients, I would still characterize most of those efforts as at least partially successful, partially participatory, and contributing toward participatory capacity building in the long-term. By building an orientation toward participatory work, using it as a means to achieve my ends and not just as a goal in itself, and looking for and seizing each emergent opportunity for a participatory emphasis as I come across it, I am thus making it the default choice for any action. My first question becomes not whether this is a situation in which I might use a participatory approach, but how can I employ some facet of a participatory orientation in the situation?

In terms of the dimension ‘levels of participation’ in Figure 6.2.1, I need to be aware that participation need not be uniform across the duration of the project for each stakeholder, and can encompass active and symbolic elements of participation by those who contribute. From my own perspective in facilitating participation in the evaluation efforts, it is never too late to invite an opportunity for participation, even if it amounts to a token effort, because the next instance may be less token, more active, and represent long-term progress.

Finally, Whitmore (1988) notes a variety of new roles associated with facilitating participatory evaluation – technical expert, facilitator, educator, and enabler:

… viewing evaluation as empowerment has clear practical implications for the role of the evaluator who must refocus attention primarily on assisting stakeholders to participate effectively. (p. 177)
The next section addresses this expansion of the roles and skills needed by evaluators working in contexts requiring participatory approaches, the ability to read and understand complex political settings, and the uneven capacities of organizations in the nonprofit sector.

Theme Eight: Roles

This section examines the changing roles of evaluators, reflecting the use of more inclusive approaches, the numerous and complex sets of relationships involved in doing evaluation in a community, the challenges of different levels of capacity within organizations, and the increasing demands for some form of accountability from multiple funders and community stakeholders. While quite a few writers have discussed the changing roles expected of evaluators (Guba & Lincoln, 1989; Mathison, 1994; Greene, 1997, 2001, 2002; Graham & Bois, 1997; Wallerstein, 1999; Ryan & DeStefano, 2001; Hanberger, 2001; Torres & Preskill, 2001; Kushner, 2002; Morabito, 2002; MacNeil, 2002; Clayson et al., 2002; Campbell et al., 2003; Rosenberg et al., 2006), this discussion has been wide-ranging, and itself reflects the diversity of roles experienced by evaluators. A key concern that I examine in the section is the contrast between how those roles are variously represented as problematic and illegitimate on one hand, or as progressive and inevitable on the other.

Evaluators working in local communities and with nonprofit and grassroots organizations can wear many hats. For one thing they can be volunteers themselves, or frontline workers in the organizations who have been drafted into the role of evaluator for their own or some other agency program. My own experience as an evaluator in such settings has encompassed being a volunteer, being a paid consultant for a program or agency, being a consultant working on behalf of a third-party intermediary agency between the funder and the delivery agency, and working directly for funders – both as a facilitator of evaluation and
as an external evaluator brought in to evaluate a program from the outside. In none of these settings was my role clear and unambiguous. Each involved negotiation, change, a multiplicity of expectations by others, and diverse actions and orientations to the job of evaluator on my part. The first vignette in this section opens the window on some of the complexity of evaluation roles when working with grassroots organizations, and builds on previous discussions of themes in evaluation practice.

**VIGNETTE 6.3.1 (GRASSROOTS) – EVALUATION VS. RESEARCH**

It is my first meeting with the new ‘team’ put together to quickly set up and implement a new front-line service initiative. The director (Laura) submitted a last minute proposal to a Ministry that had some resources available at the end of the fiscal year, and has secured bare-bones funding for a one-year pilot project for a service that she feels would complement the other two programs currently being run as demonstrations through her organization. Each of the other projects has a different funder, or rather, funders, as each has had to establish both national and provincial level funding partners. This one has only the provincial Ministry as a funder, although as usual, it requires significant ‘in-kind’ contribution by the agency and community – meaning office space, all administration, and evaluation. The latter is required by the funder, but not resourced through the project funding. Laura has allocated $1,500.00 to cover the cost of having me work with the project team to develop a plan for evaluation, support the data collection efforts, and to write up the brief findings as an adjunct to the evaluation reports I will be writing for her other projects. Since I understand the project and how it fits into the broader agency goals, I am confident that I can provide value for their money, particularly if I can work with them as a coordinator of their participatory evaluation efforts.
I am visiting the organization’s offices for another meeting, and so have coordinated the timing so that I can also meet the new project team of two part-time coordinators and three volunteers – two of whom had previously been agency clients. I know one individual from one of the other projects, but am meeting most of the team for the first time. As I walk in, Laura introduces me as “our evaluator,” but quickly backtracks on seeing the look of horror on the faces of several of the people in the room. With a concerned look, Laura continues, “but he’s ok!” This elicits a look of scepticism on the part of two individuals, and curiosity from the others. We proceed to have a discussion about what evaluators and evaluation are all about, and how that differs from simply being a “researcher”.

Over the course of two meetings to kick off my participation in the project, two broad themes relevant to my role emerge in our discussion. One is a question asked by a coordinator – if I am an evaluator, why am I getting involved now, rather than at the end? The other concerns my participation in some data collection efforts – notably a focus group with clients participating in a support group setting. A volunteer describes the issue: “These people have been evaluated continually, for years, and they have very negative associations with the term evaluation. You call yourself an evaluator, and they won’t want to talk to you. We need to just call you a researcher – that’s something they won’t be too upset by, and in fact they might even find it intriguing.”

Concerning the first issue, we discuss how short the time-frame is for setting up and implementing the project, and that I will be initially working with them to help trouble-shoot their implementation issues, and make sure they get quick feedback about how well the processes are working. I will also be helping to set up the evaluation activities that will provide the substance of the ‘lessons learned’ and final reflection on the project’s
achievements. My familiarity with the other agency programs allows me to appreciate possible and necessary links to their activities, and ways that the project can build on pre-existing processes, community linkages, resources and talents within the agency.

Someone raises the issue of deception; does it make a difference if we call it research when we’re actually doing evaluation? Consensus appears to be that it depends on what is done with the data – if this is evaluation for accountability, then respondents ought to know what the implications of their participation in the research should entail. If the results are to be used to make decisions about the extension or future funding of the project, people have a right to know that their answers might have influence for such decisions. The official vision of evaluation offered by the funder is that it should inform future program design and work done in other communities, and is not focused on accountability, *per se*. Further, the funder’s policy is to not fund extensions of projects – so this is clearly a one-shot, one-year implementation, unless funding can be obtained from another source. Even so, I remain aware that funders in general are notorious for making last-minute changes to their expectations regarding evaluation and reporting, and that even in a brief project such as this, we might have several different funder contacts as people change jobs, or take secondments or leaves. I say that I’m more comfortable calling it ‘research’ if I can then describe that research in evaluation terms – that the research is about how well the project is working, how to improve it, and how others can learn from what is done here. Such an approach engages peoples’ connections to other communities and projects, and acknowledges the potential accountability emphasis if that becomes relevant, but it also helps to reframe and reinforce the research as focused on the project, and not on individuals, as might be the inaccurate assumption that arises if clients hearing about an ‘evaluator’ in their midst.
Most people working in small nonprofit and grassroots organizations wear many hats. They change their hats daily: between meetings, within meetings, and perhaps within the course of a single conversation. They wear more than one hat at a time. Sometimes they think they are wearing one hat, when the person they are talking to thinks they are wearing another. The look, feel and size of hats, who gets to wear them, and what they mean are constantly negotiated. As people wear a hat for an extended period of time, they change it, colour it, add feathers and beads – they make it their own. Such processes also occur in larger organizations, but they tend to be slower, less transitory, more circumspect, and less subject to contradictions. Evaluators in such settings experience the full range of roles and their shifting requirements and expectations.

In the situation described in the vignette I was taking on the role of an evaluator as a collaborator – establishing a long-term partnership relationship with the organization similar to what Mathison (1994) describes as becoming a collaborating participant in the evaluation process. Mathison (1994, p. 299) suggests that such partnerships could provide a long-term relationship between evaluator and organization at less cost than an internal evaluator, and with greater likelihood of addressing systemic factors that directly affect the quality of services provided by an organization. Part of the rationale for this is that gaining the ability to examine a group of projects over time allows the evaluator to examine them more fully in their rich contexts – seeing the diversity of community and internal linkages and resources the program has available. In short, Mathison suggests,

To recast the evaluative gaze … requires recasting the role of the evaluator. One-shot program evaluations will not provide the information to address fundamental organizational traits and characteristics which influence all programs. (p. 304)
In this setting I became a key link for the project to the other programs offered by the agency. Because of my work with other agency programs, I was able to suggest community linkages relevant to their goals, and points of congruence between this project and the others being delivered. Perhaps more importantly, I was aware of the history of the agency, how it had grown, and how this specific project connected to and grew out of several previous iterations that had been delivered over the previous five years, albeit with slightly different mandates and activities. In this sense I represented part of the corporate memory for the agency, as only two of seven staff had been with the organization over this whole time period, and one of those was the director. While I had not been ‘with the agency’ for this whole period, I had interviewed many of the individuals who had participated as staff, volunteers or clients for previous programs.

My initial roles, besides corporate memory, collaborator and researcher / evaluator, also involved supporting project design and implementation. In answer to one of the first questions asked by a staff person, I was involved at the start of the project partly to help establish what would be evaluated and how, but also to ensure that the project would be evaluable. In this sense I focused on clarifying what the project was about – not simply articulating goals, but understanding what actions were involved, how those actions related to what the agency hoped would be achieved for clients, and what the agency and staff wanted to learn from the whole endeavour. As the project was innovatory, and had been put together rapidly and at the last minute, we had no doubts that it would undergo changes as staff and volunteers worked out the kinks in implementation. My hope was that I could help identify some of those kinks before they happened, build in feedback to quickly identify kinks we did not anticipate, and in short, help give the project a chance to achieve what it might.
The second question addressed in the vignette – the implications of being a researcher rather than an evaluator – reflects some of the complexities in balancing roles and wearing multiple hats. In a Canadian case study, Graham and Bois (1997) discuss this complexity by first noting the sheer number of roles taken on by various stakeholders in a community-based evaluation of a multi-funded program. In the vignette, part of the reaction to my introduction is based on the perception that an evaluator is necessarily someone who is there to check up on the program – that accountability is the evaluator’s key concern. Being hired to facilitate the evaluation portion of a project, and anticipating writing a report to funders, I had to be prepared for this possibility – first as a reaction by staff and volunteers and second by recognizing that the funder might try to insert an accountability focus at some point in the project. The concerns expressed by the agency’s staff were valid – I couldn’t just put them off with unrealistic statements that our focus was purely on learning. In the past and in other settings they had seen evaluation that focused on accountability, and so had the agency’s volunteers and clients. Rather than deny the possibility, my focus was on mitigating it.

By being a part of the project from the beginning, I was also making sure that staff understood that I knew their project, and could competently describe it. I also was walking that line between having two concurrent and potentially conflicting allegiances – to the program and to the funder. English and Kaleveld (2003, p. 40) discuss this issue, noting that program staff can perceive evaluators with trepidation, much like a tax auditor or a police officer. In part this reflects the contradiction that the evaluator is trying to balance diverse roles of being an independent reviewer while also ‘on the payroll.’ The funder for the project demanded evaluation, but did not fund it. My official client was the agency. But part of my work was to know and anticipate what would satisfy the funder, and to comply if possible.
In this way I ended up taking on a mediating role. I was facilitating the evaluation process by working with staff and volunteers who undertook evaluation activities as part of a participatory approach that devolved some of the work to them – including planning and strategizing. I was also educating these staff and volunteers about evaluation and research, and giving them tools and skills that they could apply on behalf of the program. Morabito (2002) discussed this process as one of using roles and strategies that can potentially increase the influence of evaluation in an organization. Although Morabito looks specifically at the roles of educator, consultant, facilitator and counsellor, his discussion of roles can be summarized as mediation – among stakeholders and interest groups – and facilitation – undertaking a variety of activities to simply keep things moving and make them happen.

My work with the agency encompassed all of these roles, but in particular, mediation was a key component of my orientation to conducting the evaluation. Even so, the mediation role was not always transparent to others. It involved anticipating potential requests by the funder, and being prepared to intervene on behalf of the program and agency to defend choices made. It meant preparing participants for the possibility of accountability demands, while promoting a view of evaluation at odds with such an approach. I needed to prepare for contingencies, try to remain consistent in how I talked about evaluation, but also remain open to the emergent and unanticipated learning that such an innovative project might offer.

One facet of the evaluator’s role highlighted by this vignette is that my role changed over time. Not only did it change, but I had multiple and concurrent roles and relationships with different people and programs across the agency. Campbell et al. (2003) describe a case study in which an evaluator’s roles changed as the program he was evaluating changed and developed, and as the needs of his client and stakeholders changed over time. The case study
(Campbell, et al., 2003, p. 468) focuses on how unstraightforward, complex, and dynamic the roles of the evaluator are, and how they merge in ‘thorny’ ways with other roles such as organizational development, advocacy and planning. In describing how the evaluator can become part of the program design team, they note:

> The purpose of the evaluation is to help develop the intervention; the evaluator is committed to improving the intervention and uses evaluative approaches to facilitate ongoing program, project, product, staff and/or organizational development. (p. 468)

Such a focus on the evaluator’s support of the program’s implementation helps both the program and the evaluation. It helps ensure that what is delivered is evaluable, and that any criteria used to judge the program’s achievements reflect intended program elements rather than deficiencies in implementation – particularly those arising from the short time frame of project development in small nonprofit and grassroots organizations. One of the reasons that developmental evaluation (Campbell et al., 2003; Patton, 1994) focuses on this process is that often the evaluation literature has in the past been critical of the potential biases represented by advocating on behalf of programs. Indeed, Westley et al. (2006) note that:

> Evaluators have traditionally been admonished to remain external, independent and objective, but complexity-based developmental evaluation recognizes that data collection is a form of action and intervention, that the act of observation changes what is observed and that the observer can never really remain outside of and external to what he observes… Moreover, even the roles shift, for the actors and innovators are simultaneously (and paradoxically) both the evaluated and the evaluators, and the developmental evaluation facilitator, by entering actively into the arena, even as an observer, becomes part of the action and therefore part of the evaluation. (p. 239n)

While Folkman and Rai (1997) also discuss the changing roles of evaluators over the course of a study, Themessl-Huber and Grutsch (2003) extend this by observing that the roles of other stakeholders do so as well. The power balances among and between stakeholders can realign, and present a shifting landscape for evaluators and anyone trying to make sense of the program and the evaluation context.
Greene’s work on advocacy in evaluation is another useful and complementary frame for sorting out the implications of my diverse, multiple and changing roles. Greene (1997) notes that:

In their work today, social program evaluators are inevitably on somebody's side and not on somebody else's side. The sides chosen by evaluators are most importantly expressed in whose questions are addressed and, therefore, what criteria are used to make judgments about program quality. (p. 25)

But Greene’s interest is not simply to acknowledge advocacy, but to orient it:

… advocacy is an inevitable part of evaluative inquiry, and indeed of all social inquiry today. The important question then becomes not, should we or should we not advocate in our role as evaluators, but rather what and whom should we advocate for? (p. 26)

For Greene (1997), the answer is that evaluators need to advocate not for clients or program partisanship, but for consistent, defensible, and rigorous approaches to conducting evaluation in order to ensure that it is valuable to all of those concerned. Such an advocacy approach puts the onus on the evaluator to take an active role in balancing competing interests, and placing the greater good above any one set of allegiances.

Taking on an advocacy role in an evaluation, or working to promote participation of all stakeholders can be affected by the evaluator’s link to the program’s funder, and thus to what Ryan and DeStefano (2001) refer to as “the ‘ruling apparatus’ … situated in such a way that he or she is responsible for balancing the many interests …” (pp. 191-192). This position in the ruling apparatus can represent challenges of bias and representation, but it can also provide links to resources and to other organizations, institutions, communities and ideas.

Even when the local organization is my client, there can be a perception that I am part of the ‘ruling apparatus.’ Yet this perception can work from both directions.

In writing about personalizing evaluation, Kushner (2002) notes the importance of articulating voices and portraying events that “commonly elude” evaluation. He notes how
evaluators often frame the problem as one of being ‘co-opted’ by either the client or by the subject’s perspectives if using inclusive approaches. Yet he also makes the case that evaluation also frequently marginalizes the powerful, and not just the powerless. When faced with either alternative, working to be inclusive of all perspectives, facilitating dialogue and mediating among perspectives can help to avoid some of the problems associated with participatory approaches. Indeed, while Kushner (2000) notes that there are typically considerable “political and contractual pressures” (p. 2) to attend to funders’ priorities, I have also found that the devolved context of evaluation can provide opportunities to proactively be inclusive of diverse perspectives. When the client is the community-level contractor, and the funder is an interested stakeholder with significant power but with less direct control over the evaluation process, efforts to be inclusive can be supported as appropriate and even encouraged. Funders can be genuinely interested in understanding the detailed nuances of what is done at the community delivery level, and can welcome any information or avenue for maintaining a link and presence in the project. Indeed, as the next vignette demonstrates, sometimes the evaluator is the only local link to what is happening in a community.

**VIGNETTE 6.3.2 (NONPROFIT) – CHOOSING SIDES … OR NOT?**

I am making a visit to one of four sites for the project in the northern community participating in the eight-community demonstration project. The community is approximately 60% First Nations, many of whom are geographically segregated within the community. This is my first visit to the community, and I am meeting with the project lead (Sally), the project steering committee (a multi-agency group with mostly First Nations representation), and the newly hired project coordinator (Sam, a First Nations man well-known in the community). Arriving on Monday, I meet with Sally, and then have several opportunities to meet with
steering committee members and Sam informally through the afternoon and into the evening. I attend a cultural event that evening, and this represents an opportunity to learn about local history and cultural issues through discussion with committee members.

The project is a community development endeavour that seeks to engage community members in developing comprehensive and sustainable approaches to addressing prevention. The project members have demonstrated a willingness and ability to establish multi-stakeholder coalitions interested in developing their working relationships. The community appears ready to run with the opportunity provided by the funding, and indeed, has found several co-funders so that they could expand on the approach initially envisioned in the community's project proposal.

In our meeting Sally expresses concern that she has been trying to reign in the enthusiasm of the committee, as they want to expand the project into many new areas, and she is concerned to keep the project ‘on track,’ and focused on the initiatives originally proposed. She describes herself as being a ‘stick in the mud’ for the other committee members, and solicits my assistance in requesting that the committee keep its focus and not lose sight of its stated goals. As we speak, I am aware of several issues; 1) she is the only non-First Nations person on the committee, 2) the enthusiasm and expansion of linkages to other activities is something we are encouraging in all communities, as a way of increasing their sustainability, 3) the approach of dovetailing project funding to make it fit with other activities is not just efficient for the community, it is a culturally endorsed approach to ‘getting things done,’ and 4) I am aware that the steering committee will already see me as an external person representing the funder, even though technically my role is working on
behalf of their project and in support of their needs. I am hoping to mitigate this perception of
my role as the funder’s agent in search of accountability.

I discuss all of these issues with Sally, both in terms of my support for the general
approach by the committee and her request for support. Sally’s relationship with the steering
committee appears solid, yet she also seems concerned about any potential future
disagreement, particularly as she appears uncomfortable with her role as the only non-First
Nations person on the committee. Little of our discussion is about the project’s evaluation.

The steering committee meeting the next morning begins with a round of
introductions, followed by committee members enthusiastically sharing ideas about how the
project will work. There does indeed appear to be broad-based support within the
community, and each of the suggestions presented about how to augment and enhance the
project appears to be realistic and well thought out, although I also can see Sally’s point
about the committee potentially underestimating how much work these ideas will involve if
they are all implemented.

Most of my time is spent listening. I ask some questions, and the committee members
clarify their vision of the project. The committee raises the issue of the project expansion
through dovetailing with other community activities, gently and affectionately teasing Sally
for her role as taskmaster and conscience to the group. I emphasize how this fits with the
broader goals of the project – supporting sustainability and community development, and
how the learning from this project will be useful to people in other communities. I also
support Sally’s concerns by sharing stories of the experiences of people in other concurrent
and prior projects that started off with a great deal of enthusiasm and then ‘burned out’ by
taking on too much at one time. I reframe the issue from one of ‘control’ to one of thinking
long-term rather than short-term, and to the committee and community maintaining their
flexibility to respond rather than being locked into many concurrent activities, some of which
may prove to be less feasible than others.

Our brief discussion about evaluation focuses on participatory approaches and
emergent research design relevant to such a grassroots endeavour. We also discuss the
broader role of the project as a demonstration, and how the implementation and the
evaluation will provide an opportunity for sharing with and learning from other communities.
We conclude by discussing my role as a resource to the community and the committee. I
frame my relationship with the funder as one of capturing and translating what the
community does – what happens, what works and what does not – not as accountability, but
as documentation and engagement of a public dialogue with funders and those in other
communities. The committee seems receptive to this approach, but their reactions remain
measured and cautious, reflecting their experience with prior projects and evaluations as
heavily focused on accountability.

Reflection 6.3.2

As in Vignette 6.3.1, this vignette portrays a situation in which I work with an
implementation team to define the boundaries of my participation in the project. Unlike the
first vignette, this one portrays a situation in which the multiple roles are less explicit, and
require ongoing efforts to identify and articulate the various nuances of my actual and
potential contribution. In part, this reflects the ambiguity inherent in my role set. In this
setting I am the only person representing the funder who will visit the community, and thus
the main link in terms of both implementation and evaluation. In turn, I report to a central
coordinator who oversees the implementation on behalf of the funder, and who, in a sense,
provides an additional buffer between me and the funder’s representative, who I have met only one time.

Although I provide the funder with verbal feedback about the status of the project, the local project leader prepares official status reports, which I receive, but do not vet. My efforts to understand and evaluate the progress of the project contribute to finding ways that the central agency could support the project, but do not represent efforts to keep track of the project. Yet my final report will go to the funder, as will any interim reports I prepare, and as such the funder’s representative is very interested in learning whatever detail about the project I may be able to provide. For some of those in the local community and on the project steering committee, this can translate to a perception that I am visiting the community to ensure compliance with funding guidelines and expectations.

Rosenberg et al. (2006) examine evaluation as a constantly renegotiated process (referencing Folkman & Rai, 1997), and as such, recommend that evaluators negotiate an expanded role beyond ‘evaluation’ – one that encompasses mediation of diverse interests. In the situation portrayed in the vignette it was necessary to negotiate all of my roles, as the various players each had a somewhat different perspective on what I could or should do in the situation. For example, the project lead viewed my visit to the community as an opportunity for her to reinforce her own leadership and legitimacy, and so she was invested in seeing my role as representing the funder, and project authority. For some of the committee members, I was a potential antagonist – representing the funder’s impending arbitrary and invasive imposition of control over the project. For the funder and agency coordinator, my very presence in the project represented a measure of control – as the only external person to visit the community I became perceived as part of the ‘ruling apparatus’
no matter what my actual actions or intentions were. Yet I also represented the community’s connection to the outside world – a window into the community, to be sure, but also a voice out of that community. While the initial orientation towards my role set by members of each stakeholder group involved trying to identify and influence my allegiances in the setting, my own efforts were to try and maintain flexibility in how I was perceived, so that I could change roles and relationships over time, but also to break through some of the preconceptions about the nature of the role of evaluator in such a setting.

In writing about some of the challenges and lack of conceptual precision in how the concept of ‘role’ is used in social science, Giddens (1984) notes such social positions are constituted as the intersection of “significance, domination and legitimation” (p. 83) of agents in social settings. A social position – better described as active ‘positioning’ according to Giddens (1984, p. 84) – encompasses expectations on the part of both the actor and those with whom he or she interacts. As such:

A social position can be regarded as a social identity that carries with it a certain range (however diffusely specified) of prerogatives and obligations that an actor who is accorded that identity… may activate or carry out: these prerogatives and obligations constitute the role prescriptions associated with that position. (p. 84)

And yet the normative rights and obligations of identities – such as evaluator as presented in the vignettes – are often not clearly formulated in contemporary society, which is typified by a multiplicity of serial and concurrent identities. Some stakeholders in the setting have a sense of what they expect of an evaluator, but many have little prior expectation. And even those with a clear idea of what the evaluation role encompasses, have different expectations than I do in numerous respects. For that matter, another evaluator in the same situation would likely harbour a different constellation of expectations regarding their rights and obligations, reflecting the unique character of each project, community and evaluation context.
The evaluator’s position is unique in the settings described in these vignettes. Other stakeholders have in-depth knowledge about their own working contexts, and indeed, this knowledge is crucial for the successful implementation of the program and conduct of its evaluation. Yet the evaluator must master both local and centralized knowledge sets, reconcile the two, and in many ways bridge the two worlds. Or rather, the evaluator must bridge the many worlds represented here, including those already mentioned, and also including the perspectives of the program’s clients, and increasingly, the academy – as an interested party to and resource for programs and services of all stripes.

The traditional evaluation emphasis on avoiding bias by resisting the call to become too closely aligned with the project reflects a bias in favour of those already most powerful in the evaluation and program administration setting. In its typical formulation, this emphasis is on not taking the perspective of the program deliverers or clients. And yet as the two vignettes have demonstrated, it is as easy to become – in effect, even if not in intent – part of the ‘ruling apparatus’ of the program. By acknowledging these competing interests, and working to balance the perspectives without succumbing to any one of them, the evaluator is creating a new type of role set. This encompasses such roles as mediation and facilitation, as discussed by Greene (1997), Folkman and Rai (1997), Hanberger (2001), Schwandt (2001a), Morabito (2002), MacNeil (2002), Clayson et al. (2002), Campbell et al. (2003), Markiewicz (2005), and Rosenberg et al. (2006), but it also appears to represent something more.

Beyond the variety of roles already noted thus far – collaborator, co-implementer, participant, advocate and mediator – writers have begun to document a wide range of roles that have been part of the evaluation position. Recently, evaluators have documented taking on such roles as counsellor (Morabito, 2002), educator (Morabito, 2002; Schwandt, 2001b &
2002), coach (MacNeil, 2002), negotiator (Rosenberg et al., 2006; Markiewicz, 2005; Owen, 1998) and interpreters, translators and storytellers (Clayson et al., 2002).

Many writers start with the basic assumption that the evaluator brings technical expertise to research activities (Graham & Bois, 1997; Minnett, 1999; Morrisey et al., 1997). And yet there has recently been a growing current of writing and discussion in the evaluation literature critical of this technical focus. This literature examines how evaluation work is grounded in and “constitutive of” the political currents and dynamics of practice. Such writers as Guba and Lincoln (1989), House (1993), House and Howe (1998), Greene (1997, 2002b), Kushner (2000) and Hanberger (2001) have reflected on the role of the evaluator as a player – as an active agent of change or of maintaining the status quo. This discussion is part of the broader debate that seeks to challenge the methodological preoccupation of the discipline, and in the words of Greene (2002b) the commonly held view of evaluation:

… as an activity practiced at a distance, apart from the contentious fray of political decision making about resources and priorities, in a space somehow shielded from special interests and advocacies, a bit like a little air bubble safe from contamination by bias… (p. 1)

This deep-seated sense of denial of the place of the evaluator in the research process – of living in the bubble – has helped to obscure the diverse roles of the evaluator. It posits evaluators’ actions as most effective when they are virtually invisible, or completely replaceable by any other competent technician. Greene’s argument is that this is a question of both what is current practice and what ought to be practice by evaluators. Greene’s work seeks to articulate and legitimate a broader conceptualization of the discipline. It fundamentally asks whose interests evaluation does and should serve, and the kinds of values, norms, knowledge claims and power relationships that are influenced by evaluation, and in turn influence its practice (Greene, 2002b). She notes:
These questions point to the intersection of evaluation with democratic theory, for the ideals of democracy provide the most inspiring and inarguably the most defensible normative agenda for evaluation. (p. 4)

While the next section will examine in more detail the issues of deliberation and democratic evaluation, the concepts also provide a backdrop for better understanding the roles of evaluators. For example, Greene (2002b) and MacNeil (2002) introduce the overarching concept of ‘stewardship’ that can help us understand and reflect on the diversity of roles in contemporary evaluation practice. They discuss stewardship as an orientation towards practice, and to taking on responsibility for the impacts and outcomes of the evaluative endeavour. MacNeil (2002) focuses on the facilitation role of the evaluator in making deliberation possible:

Evaluators as stewards of citizen deliberation can choose to make an investment in social capital by attending to the social and perceptual connectedness among stakeholders participating in evaluations. (p. 46)

MacNeil’s emphasis includes stakeholders who are not immediately present in the context of evaluation activities. In the vignettes discussed in this section, this can also include those who are not (yet) aware of the evaluation, or of its relevance to their lives. As a steward of citizen deliberation, the evaluator also works to identify relevant stakeholders, and find appropriate means to include them.

In the context of devolved programming, such efforts to promote and facilitate deliberation appear to be particularly vital. As documented in previous sections, the trends toward decentralization and devolution can disconnect communities – a process facilitated in part by the devolution of evaluation, given that broader program-wide evaluations are often replaced with narrower organization or community-based ones. Finding ways to build and maintain connections within and between communities requires positive effort to compensate for this narrowing of the range and content of deliberation, and evaluators can be in a
position to facilitate such efforts. In the vignettes documented in this section, my work as an evaluator served to provide some of those links to other organizations and stakeholders within the organization and community (Vignette 6.3.1) and to other communities (Vignette 6.3.2). By facilitating deliberation and connection between groups and communities, the evaluator widens the types of evidence used in conversations about the value and benefit of programs, and can establish or build a public dialogue that benefits the broader policy environment, but also helps individual programs, agencies and communities by linking them to that debate. And as Morrisey et al. (1997) and Love (2004) point out, a significant role for evaluators is linking evaluation practice with the academy. Morrisey et al. (1997), discussing challenges faced by community-driven evaluation initiatives, address bridges between prevention practice and prevention science and roles evaluators can take, included providing technical assistance on accessing literature, encouraging the use of appropriate resources, and facilitating the transfer of information back and forth between science and practice. Love’s (2004) emphasis is on knowledge mobilization – again building upon the knowledge relevant to a community’s activities, but also connecting the community and ensuring that the knowledge developed is not lost, but is able to inform beyond individual contexts.

Clayson et al. (2002) note another facet of connecting communities – that if evaluations ignore the broader context of social issues that precipitate programs and services and focus only on local solutions and remedies within a community, they will inevitably neglect some significant elements of program rationale and the framework within which solutions are constructed. Sometimes both causes and solutions lie elsewhere, beyond the means of those to whom responsibility for remedy has been delegated within a community. This broadening of the discourse helps to identify common problems and issues, but also
orients solutions to systemic factors rather than limiting them to immediate and individualized ones. Clayson, et al. (2002) argue that failing to address such issues can render the evaluation culturally and contextually incomplete:

… because evaluation is a politically contextualized act, all aspects of a particular evaluation, including its design, implementation, outcomes, and uses are shaped by the power relationships among the stakeholders. All the stakeholders in this process are in some respects vulnerable. (p. 42)

Negotiation within evaluation practice is a relatively new idea (Owen, 1998), and yet it appears to hold significant potential as an evaluation activity in support of mediation, facilitation and leadership roles – of addressing the vulnerability of stakeholders within broader political contexts that impose evaluation requirements. Recent work in the planning field seems particularly relevant, as it tries to develop analytical tools that can help capture the complexity of negotiation (Sork, 1996, p. 84). Referencing the work of Elgström and Riis (1992), Sork examines some of the implications of using negotiation theory and frame-factor theory to understand and illuminate the character of negotiation in planning practice.

In a study of the Swedish curriculum revision process, Elgström and Riis (1992) combined negotiation theory and frame factor theory and reflected on how frame factors affect negotiation processes. Frame-factors “constraining the intellectual space and the space for action within a process, which the actors at each point of time during the process cannot influence or perceive they cannot influence in the short term” (p. 104). They further note that “frame-factors may be perceived as banning or preventing certain alternatives and solutions. Such elements may then be perceived as non-negotiable” (p. 104, italics in original). This perception becomes a key element of how frame-factors are relevant to negotiation in evaluation practice.
Elgström and Riis make a number of distinctions about negotiation, of which one seems particularly relevant to this discussion. They distinguish between distributive and integrative bargaining:

Distributive bargaining is connected with a competitive orientation among the negotiators. The process contains high initial demands, threats, manipulations and a win-lose approach. Integrative bargaining denotes search processes involving creative problem-solving. The search is for new options with the hope for mutual gain. (p. 102)

In combining negotiation with frame-factors, they note that frame-factors can themselves sometimes become the object of negotiations:

Such attempts to change the existing frames may be labelled *meta-negotiations*, to be differentiated from the *substantive* negotiations which take place under the umbrella of existing frame-factors. Negotiations may occur *within* given frames, but also *about* these frames….

In any decision-making process, some frame-factors are by all actors perceived as given and impossible to change. Others are, at least by some actors, considered to be negotiable. (pp. 104-105, italics in original).

Elgström and Riis note that meta-negotiation is more likely to occur in non-specified situations, when actors are in difficult negotiation positions, or when the negotiations become deadlocked (p. 105).

For Sork, the perception of whether frame-factors exist is a key part of this process. Believing that a factor is part of the frame and thus beyond negotiation can prohibit potentially successful action, while trying to negotiate material constraints as if they were freely negotiable can be both frustrating and interfere with planning processes (p. 84).

As I apply these ideas to my work evaluating and negotiating with nonprofit organizations, I observe that smaller organizations seem to perceive most factors as frame-factors and beyond their control. Part of my work with such organizations is to help them to identify which factors are part of the frame, and which are negotiable. An example is the case
described in Vignette 5.1.1 in which I called the provincial contact to clarify the expectation concerning the use of logic models in the proposal being developed. Beyond this, some may also be perceived by the funder as frame-factors, but become open to meta-negotiation based on reasonable requests on behalf of a local client. For example, in Vignette 5.3.1 when the federal funder’s contact came to understand some of the unintended consequences of developing and introducing a new database, I was able to somewhat mitigate those impacts by reconciling administrative goals with local practicality and politics.

I have seen cases in which larger nonprofit organizations have experienced some success in establishing meta-negotiation about some frame-factors. I have also seen such organizations experience frustration as they come to assume that perhaps all factors are thus negotiable, and become stymied when they find that some frame-factors are indeed immutable. The larger lesson in such situations is that there are often many ways that a factor may be a frame for some actors, but be open to some negotiation by others. I certainly experienced this when working as an internal evaluator in government, watching as others made policy and program decisions with minimal reference to my evaluation efforts – there were factors not only beyond my ability to influence, but also beyond my recognition.

Nonprofit organizations similarly may see funding or policy decisions made concerning programs that they are delivering, but not understand how their own evaluation or reporting activities contribute to those decisions. An evaluator in a mediating role among funders and deliverers may be in a better position to understand some of that decision-making process – I certainly understand them better now than I did when I was immersed in the situation within government. For my local community clients, this can clarify
unintelligible processes in a way that can empower action, and remove some of the frustration that comes from what may be perceived as bad-faith negotiations by partners.

Looking at the meta-negotiation of frame-factors in this way can be a useful means of thinking about how evaluators can provide day-to-day leadership. As individuals who occupy a bridging niche among and between nonprofit delivery agencies, funders and other stakeholders, evaluators can represent one to another, interpret, negotiate, reframe and generally facilitate a host of important evaluative and implementation factors.

In their reflections on evaluation as stewardship, Greene (2002b) and MacNeil (2002) similarly appear to be talking about aspects of leadership within evaluation practice. This vision of stewardship and leadership recognizes that evaluators work in contexts in which they have some power, but also need to be attuned to the subtleties and varieties of power among stakeholders. It also reflects the ways that leadership in evaluation contexts – particularly the kinds of informal and diversely structured contexts found in small nonprofit and grassroots organizations – is better understood as types of action within these contexts rather than a specific role to be taken on. To ground the discussion of roles in evaluation practice, and understand the bridging between the role and the activity of leadership it is useful to introduce and explore the concept of skill.

Much of the recent work examining dialogue and democratic evaluation ends up exploring evaluation skills, but in a roundabout way. When House and Howe (2000), MacNeil (2002) and others discuss dialogue and democratic evaluation, they typically begin by referring to them as outcomes to be achieved, rather than actions to be undertaken. Even Greene (2002b), in describing the discipline’s tendency to have evaluators assume a position
on the sidelines or in a protected air bubble, uses wording that emphasizes the discipline and
the practice of evaluation, rather than actions by evaluators or people doing evaluation work.

... evaluative results cannot be generated from nor represent a neutral position
within the politics of the contexts in which we work. So, like it or not, the practice
of evaluation itself either sanctions and reinforces, or alternatively challenges and
disrupts, key dimensions of these contexts... (p. 3, italics in original)

Most discussions of skill in the discipline have typically focused on technical
evaluation skills (Caron, 1993), although recently King et al. (2001) and Stevahn et al. (2005)
have explored a wider variety of critical ‘reflexive’ skills. In part, this reflects Greene’s point
about evaluators working ‘in the bubble,’ but it also reflects the diversity of contexts in
which evaluation work is done, and the multidisciplinary routes to evaluation practice taken
by evaluators. The technical skills are usually described in such a way that they appear
consistent, transferable, measurable, teachable, and independent of context. By contrast,
leadership or dialogical evaluation skills are context-dependent, transient, hard to define and
perhaps even harder to ‘teach.’ While technical skills are typically obtained through some
combination of training, education and experience, leadership skills usually represent a much
more significant investment in experience and embedded knowledge.

When Hanberger (2001) suggests that evaluators act as mediators, he offers a caveat
to what this might mean for the changing role and skill set of the evaluator:

Taking evaluation to the people this way also requires different skills and
competence to act as a mediator. Most evaluators are not trained to take a key role
in deliberative processes, and they would need some theoretical orientation and
practical guidelines.

This focus on the skills and attributes of the evaluator is reflected in Kushner’s examination
of the activities and relationships appropriate for democratic engagement – being
circumspect, respectful, safe but productive of public insight, respectful of respondent
autonomy, and strategically intimate (2002, pp. 6-8). Ryan and DeStefano (2001) also
address such skills as they attempt to reconcile how dialogue broadens the range of activities of evaluators beyond the discipline’s traditional narrow focus on research methods:

We find dialogue in evaluation to be rich and perplexing… Power relations, the identity of the evaluator, and the skills needed to facilitate dialogue are problematic, including virtual dialogue. The evaluator needs to be more than an expert. (p. 199)

Ryan and DeStefano focus on skills, and particularly ‘tacit’ skills as “a particular kind of expertise developed … in the ‘thick of practice’ by the evaluator” (p. 192). These skills include sensitivity to the nature of relationships within and between stakeholder groups and between evaluator and stakeholder groups, mediation, conflict negotiation, and interpersonal skills. Greene (2001, p. 183) also focuses on dialogue in evaluation as a context that demands a new set of skills. For Greene, dialogue in evaluation needs to be both respectful and equitable – involving “everyone as speakers and listeners and learners.”

A sampling of the variety of skills now being acknowledged as relevant to evaluators includes mentoring (Minnett, 1999), social or interpersonal skills (Minnett, 1999; Ryan & DeStefano, 2001; Greene, 2001), mediation (Hanberger, 2001; Ryan & DeStefano, 2001; Greene, 2001; Clayson et al., 2002), negotiation (Markiewicz, 2005; Rosenberg, 2006), conflict management (Ryan & DeStefano, 2001; Markiewicz, 2005), facilitation (Torres & Preskill, 2001; MacNeil, 2002), cultural and linguistic competency (Clayson et al., 2002), and political astuteness (Ryan & DeStefano, 2001; Greene, 2001 & 2002b; Hanberger, 2001; MacNeil, 2002; Clayson et al., 2002; Kushner, 2002; Campbell et al., 2003; Markiewicz, 2005). The latter can encompass knowing who is at the table, who needs to be there, and how to get people a seat at the table. It can involve recognizing when there is a problem, having the sensitivity to appreciate what is said and what is not, and understanding the dynamics of relationships among stakeholders and between stakeholders and evaluator. It involves building and maintaining trust among groups with diverse interests and priorities over a
period of time. And it involves the skills to go beyond seeing problems and recognizing what is happening, to being able to do something about them. This balance between understanding and action can be encompassed by the idea of authenticity.

Terry (1997) and Evans (2000) discuss authenticity by focusing on the nature of trust and the need for it to be related to worldly competence and capability, which Evans calls ‘savvy’ (2000, pp. 287-288). Both writers contrast authenticity with sincerity (Terry, 1997, p. 109; Evans, 2000, p. 289): we may believe someone, and even believe that they mean what they say, but not trust them because we don’t believe they can do what they say. Without the capability of following through, it is difficult to engender trust in others. This confluence between trust and savvy is relevant for evaluation, particularly in nonprofit organizations.

As noted in Vignette 6.3.2, the evaluator may be the main point of contact between funder and deliverer, and thus have an important role in accurately and fairly representing each to the other. Phillips and Levasseur (2004) describe this context by noting that with the devolution of programs from government to nonprofit organizations, many of the people who have previously delivered programs within government have been moving into more administrative types of work – often overseeing the delivery of programs by those working in the nonprofit sector. Over time, these administrators become less aware of what is happening with clients, and potentially out of touch with the working reality of front line program deliverers. So those working on the front lines of nonprofit organizations can lose trust in the people who are their links to funding, program design and development, and accountability. Administrators can lose the ‘savvy’ that has in the past given them authenticity in negotiating and developing contracts and designing program delivery in local communities, and in demonstrating leadership within the programming community.
In this way, the evaluator is expected to understand both the corporate world of the funder and the nuts and bolts of local implementation. This provides both an opportunity and a responsibility to be a responsive, trustworthy and ‘savvy’ face of each with the other, and to help each develop and build the trust and savvy on which their long-term relationships will be based. Certainly the example from Vignette 6.3.2 demonstrates a comparable situation. In order to accomplish the evaluation work required of the project, I needed to also undertake a variety of project and community development activities involving a diversity of roles and non-technical skills. This work was a necessary condition for the success of the evaluation, but it was also a necessary condition for the project’s success as well.

What roles did I use in the settings described in the vignettes? By assisting with the implementation of each project, I was a collaborator over a period of time, and in some ways, a participant in processes and activities initiated by the projects. For the project described in Vignette 6.3.2 I took on roles that involved mediation (between the steering committee and the project lead, and between the funder and the community stakeholders), and advocacy (in Greene’s [1997] terms, on behalf of a solid implementation, and research goals aimed at both the local community and the broader society). I also told the story (translated, interpreted) of the project to the funder, and portrayed the funder’s story (and expectations, hopes, values) to the local stakeholders – representing each to the other in a way that authentically allowed individuals to see beyond the stereotypes and bureaucratic expectations to the real people who were connected through the project, but might never meet in person. For both projects I served as a mentor and coach and educator about evaluation, but also about the processes of working together, problem solving, building consensus, and being inclusive. In short, I strove to build a sense of the projects that went beyond the immediate contexts by connecting
several communities and establishing aspirations for both projects that might ensure long-
term efforts that could withstand turnover, resource challenges, unanticipated changes, and
the vagaries and messiness of community development and research on the fly.

What skills did I use in undertaking these roles? I listened carefully to all of the
stakeholders, asking questions and working to build an environment of trust and inclusive
approaches to discussion and action. I spoke in culturally appropriate ways – at times when it
made sense to do so, and in a way that supported the culture of the environment (both
projects involved working with and for First Nations communities in a multicultural context).
I facilitated and mediated and negotiated based on need and context – building my efforts on
being able to read the requirements of the situation (political, cultural, ethical), while
balancing the twin priorities of consistency of messages with flexibility in responsiveness. I
worked hard not to take sides by actively taking all sides. I used humour and play (Schwandt,
2001b) and took risks (MacNeil, 2002) in order to engage while establishing a connection
among other people, and building ownership for the outcomes by those undertaking them.
Finally, I used technical knowledge to build realistic plans and connect local efforts to those
in other communities, including the academy.

Was this stewardship? My actions went far beyond the kind of work that I had ever
studied in any course on how to do evaluation. I used skills and knowledge obtained through
experience – and indeed, much of that experience was in the practice of doing these projects.
The ‘tacit’ knowledge of what to do grew out of my reaction to the problems and challenges I
experienced as I was doing it, and reflected the uniqueness of each of those settings in
making it imperative that I figure out what to do. Few official evaluation resources offer
advice about such contexts. The key sources of my improvisations were discussions with
other evaluators, colleagues, and also those within the communities. My resources were primarily informal, even though I worked to question my actions in a way that allowed me to learn from the situation and capture the process for future reference.

Was this leadership? Certainly it was a context in which leadership was called for, and for which the official assignment of leadership roles was ambiguous, at best. By taking on the task of connecting people beyond the immediate context, and using a transformative emphasis on inclusiveness and bridging settings, interests, and attending to inequalities in power and access to resources, I was going beyond traditional lines of authority to build consensus and promote collective action. It was not leadership attached to a clear role structure, but to me it represented leadership in action. If I had not undertaken the role – even in this informal way – it likely would not have been taken on by others in the setting. And if I had listened to the traditional emphases of the evaluation profession, I would likely have avoided such a role as representing a loss of independence, objectivity, or impartiality.

Some of the valuation that we attribute to stewardship and leadership skills in evaluation comes from their perceived scarcity – perhaps based on the reluctance evaluators have for taking on such roles. We certainly are more likely to notice leadership when it is missing or inadequate than when it is present or effective. Attewell (1990) notes that one of the ironies of skill in general is that the more skilled we are at something, the less apparent the skill may be; as we gain facility, we somatize the skills and they become unconscious, and that is a reflection of our competence. So we are perhaps less able to articulate and describe skills that we know well, as they are virtually automatic to us. For leadership situations, this fluidity of skill – that we are able to act without drawing attention to the actions and to the skills – is a key element to its success. The more we are seen to be trying to
lead, the harder it can be to do, and the more it comes across as controlling behaviour rather than leadership. This also contributes to the difficulty in acknowledging and identifying leadership, stewardship and other non-technical skills as relevant to evaluation practice. It also helps us to understand why Ryan and DeStefano refer to these as ‘tacit’ skills.

The socialization process usually envisioned for evaluators emphasizes learning how to tread carefully in order to remain ‘in the bubble,’ to use Greene’s term. So how do we develop these skills ‘in the thick of practice’? Do we leave it to chance? Do we keep these skills a deep, dark secret, well hidden in the evaluation closet, or perhaps in that hidden drawer of the proverbial evaluation toolkit? I believe that discussing and embracing the role of leadership in evaluation practice will enrich our practice, and help us to identify and develop these skills openly, rather than as the main point of stories of glory and horror told in the corridors of conferences. Ultimately, they will benefit the clients and stakeholders of evaluation, who just might find evaluation to be more palatable, relevant, and useful.

This section has explored some of the ways that the roles of evaluators have changed in recent years, and how these new roles are particularly relevant to the context of devolved evaluation practice. The next section builds on this reflection about new evaluator roles, and specifically links the ideas to the evaluator’s efforts to promote dialogue, and apply ideals of deliberation and democracy in evaluation.

**Theme Nine: Dialogue, Deliberation & Democracy**

The evaluation literature has recently begun to examine the role of dialogue in evaluation. Some of this work has built upon the work of Burbules (1993) and his examination of dialogue in teaching. Burbules’ starting point is three general issues of postmodernism – the centrality of an analysis of power, the emphasis on the irreducible
plurality of cultural world views, and the quest for alternatives to the intellectual goal of finding underlying principles, generalizable rules, and universal definitions (pp. 2-3).

Burbules contrasts teleological and non-teleological views of dialogue. Teleological dialogue aims to use a dialectical process to uncover ‘truth’ – the assumed goal of participants in the dialogue. Non-teleological dialogue, on the other hand, aims to have participants in the dialogue progress towards greater approximations towards both a common and an individual understanding – something that extends beyond what could be achieved through individual effort.

Burbules is trying to move dialogue beyond a zero-sum conception of truth or understanding. In this vision of dialogue, different versions or approximations to truth do not necessarily need to be in conflict. While those engaged in dialogue may develop shared understandings, they also can assist one another in constructing and refining individual perspectives and understanding, while ultimately agreeing to disagree (or perhaps, not even actually recognizing the other’s understanding). In this way, dialogue may involve both inter-subjective progress and interactive construction of individual understanding.

This conception of non-teleological dialogue can be seen as “a developmental communicative interchange through which we stand to gain a fuller apprehension of the world, ourselves, and one another” (p. 8). Following Friere, Burbules emphasizes three critical aspects of dialogue: its relational character, its constructivist view of knowledge, and a non-authoritarian view of teaching. One further aspect of Burbules’ view of dialogue relevant to evaluation is that it can encompass group contexts, and not just dialogue between two people. In this way participants learn to express themselves to others, learn to listen, and learn about other people and ideas (p. 9). In sum, Burbules (1993) notes,
In pedagogical encounters, we do not change other people. They change themselves: They construct their own understandings, they change their minds, they decide on alternative courses of action, they redefine their priorities, and so on. (p. 10)

Those who have worked to incorporate dialogue into evaluation have also described different types of dialogue. Ryan and DeStefano (2001) distinguish between concept – an inclusive and divergent attitude of tolerance and cooperation aimed at mutual understanding, and consensus-building dialogue – an inclusive and convergent dialogical process involving cultivating tolerance and respect for diverse perspectives (p. 191).

Karlsson (2001) references Bohm (1996) in distinguishing three forms of discourse: everyday (ordinary) conversation – “spontaneous movement between asking and answering questions about each other’s lives,” discussion – “an exchange of opinions in a negotiation context,” emphasizing limited analysis, and dialogue – “an exchange of ideas and meanings” with the aim of learning (p. 212). Karlsson compares the latter with Buber’s (1961) idea of genuine dialogue – “an exchange of ideas and meanings that develops our thoughts and promotes awareness of our thoughts and values. In a dialogue nobody is trying to win; and everybody wins if nobody wins” (p. 212).

In practice, contexts in which dialogue is most relevant to evaluation are also ones that require instrumental decisions based on such deliberation. While instrumental decisions may be immediate and inexorable, the background dialogue involving values and broader awareness of issues and relative positions often remains an ongoing concern. Specific decision points represent occasions to revisit such dialogue, beginning with discussion and debate, but often eliciting opportunities to address new perspectives on central issues. The following vignette describes a setting in which the need to undertake immediate and concrete decisions about project and evaluation implementation represented both the need for and chance to establish community dialogue.
The Northern community in the eight-community pilot project has been experiencing some challenges. The individual who established the local project – wrote the proposal, worked to build the local inter-agency coalition overseeing the project, and chaired the committee for over a year – has left the project and the community. The on-site coordinator, Virginia, has been frustrated by the apparent leadership vacuum within the committee. For administrative purposes, she was hired by one of the participating agencies, but reports to the inter-agency committee as a whole. The newly nominated committee chair has been focused on other concerns, and does not appear comfortable with the role of overseeing the work of the coordinator. Virginia’s efforts to seek guidance from the person in her employing agency to whom she nominally reports have elicited what seems to be a deluge of demands on her time by various committee members, many of which are apparently contradictory. In addition, the director of her employing agency has been pressuring her to undertake some agency activities, even though all of her time is supposed to be committed to the inter-agency coalition project.

Virginia has turned to me for advice and support. She tells me that she feels like she has “twenty bosses, each of whom has a different set of activities and priorities for me to undertake.” Some committee members have approached her with requests that seem to contradict the broad direction she has received from the committee, as well as specific requests from other committee members. She feels she has reached a state in which she feels paralysed into non-action.

As the off-site evaluation support to the project, I have limited opportunity or authority to intervene, despite my role as the primary face of the funder in the community. I
emphasize to Virginia her own ability to assume a leadership role in how she interacts with the committee. Rather than taking on the requirement of needing to choose between conflicting requests among committee members, I suggest that she bring the issues to the committee for decision, and let the committee take on that task of deliberation in an open and public space. She can provide background information to help the committee understand the issues and the decisions needed, and encourage the committee members who bring her specific requests to use the committee process to introduce alternative interpretations or actions. I emphasize that this does not mean telling the committee what they must do, but helping them be clear about what their decision points are, and what their own roles encompass. I compare it to the role of an Executive Director who reports to a Board of Directors, and goes into meetings knowing there will be disagreement about issues, but also aware that the disagreement resides outside her own responsibilities. The director’s role is to facilitate resolving the immediate issue, not to take on responsibility for finding every possible solution or even to establish committee agreement about the issue. I also suggest that one way to facilitate this deliberation is to identify and articulate the ways that she sees each issue affecting those stakeholders who are not at the meeting – the clients, agencies and others who are affected by it but who are not present to represent their own concerns.

I arrange to participate at some of the committee meetings via a speakerphone connection. My first emphasis is on helping the coordinator to clarify reporting lines and the role of the committee and coordinator. I use a comparable situation in another community participating in the project as an example, describing how they resolved the issue, and suggesting that there may be other solutions that make sense for this community. I believe I am helped in this situation in that although there are differences among committee members,
they do not seem to be split into factions, and remain receptive to learning about the other ideas at the table. Although the committee chair remains relatively uninvolved, the coordinator and I are able to initiate some good discussion about priorities and resolving next steps for her and the committee, as well as surfacing several long-standing issues that seems part of an ongoing discussion familiar to many of those around the table. The committee members with alternative ideas about what to do seem happy to have been heard, although they ‘agree to disagree’ on the immediate issues at hand. They acquiesce for the moment, with the understanding that the issues will be revisited at subsequent meetings. Virginia has a concrete plan of action that she feels will carry her through until the next meeting, and the committee has a process for raising their concerns.

**Reflection 6.4.1**

Schwandt (2001a) identifies some of the potential ‘outcomes’ of dialogue as including:

(a) not agreement, but a common understanding in which parties do not agree but establish common meanings in which to discuss their differences; (b) not a common understanding, but an understanding of differences, in which the parties do not entirely bridge these differences, but through analogies of experience or other indirect translations can understand, at least in part, each other’s position; (c) little understanding, but a respect across differences, in which parties do not fully understand one another, but by each seeing that the other has a thoughtful, conscientious position, they can come to appreciate and respect even positions they disagree with. (p. 232)

My work with the community will encompass perhaps a year and a half duration. Most of the people on the inter-agency committee have known one another for years, have worked together on different committees and task groups many times, and may be quite familiar with one another’s views and concerns. It would be very difficult for me to try to appreciate all of the subtleties of their perspectives within the time available to me, and indeed, perhaps there is little need, other than that some of them impinge on a few of my
immediate instrumental concerns. But I do recognize that the specific issues we address tie into long-standing discussions and perhaps grievances. Providing an opportunity for those to be aired serves several functions. It surfaces differences and the rationale behind them for those new to some of these issues. It enables participants to learn about the positions and meanings of others in a potentially new way – perhaps extending the long-term dialogue into novel territory. It also reassures people that they can differ on important matters – but that life goes on, and by respectfully addressing those differences once again, they can develop or hold onto the belief that someday they will be agreed with or understood, and at the very least, their input and stance are respected enough to be heard.

Part of my contribution to this process is simply legitimating the value of the dialogue, and that resolving the immediate issues is not the same as resolving the long-term differences, which may never happen. By recognizing that differences exist, encouraging and perhaps facilitating discussion about them, and then bringing in additional resources from outside the community, the evaluator can sometimes change the character of the discussion, and facilitate progress when things might seem at an impasse. Such resources might be in the form of anecdotes or connections to other communities coping with the same experiences, or resources from the literature or the academy. More often than not, the resources identified are already present in the community, and the evaluator’s role can be to simply identify them and label them as reasonable for inclusion in the discussion.

A complicating factor for working with smaller grassroots or nonprofit organizations or community coalitions is the limited time available to spend in the community and get to know the local issues and players. This is exacerbated by the typically short duration of projects, and can be somewhat mitigated by having the opportunity to participate in several
projects over a period of time. It can also be challenged by needing to work with and through intermediaries such as the project coordinator, who must cope with her own barriers to participating in the project and evaluation, including multiple roles, overlapping loyalties, ambiguous and shifting power dynamics within the setting, and sometimes, as in this vignette, the reality that the ostensible employer is not the boss – although they perhaps were last month, or might be tomorrow.

MacNeil (2002, pp. 47-48) discusses deliberative forums as a means of constructing spaces where people can participate in the exchange of values through dialogue. She notes how critical it is in constructing successful deliberation for the evaluator to have a good understanding of the power relations operating in the situation. This can be problematic because many of the power relationships are tacit and are not immediately obvious, especially to an outsider. While MacNeil suggests extended participant observation as a method for understanding and dealing with perceptions about the power distribution among stakeholders, such an intensive process can represent a considerable challenge for evaluators in devolved settings, as it necessarily entails spending a fair bit of time among people who may have significant power within the context but little immediate role in the evaluation process. Gaining access to such stakeholders, particularly when the evaluator is not resident in the local community, can represent a significant burden. MacNeil starts by conducting individual interviews with all of the vested stakeholders that she can, as early in the process as possible. She follows this by including those with less authority in the dialogue as early as possible and making sure that they receive the same information as everyone else involved.

In examining some of the challenges safeguarding privacy and confidentiality in a deliberative forum, MacNeil (2002, pp. 49-50) notes that while the evaluator may take
careful precautions in safeguarding information in private discussions, other participants in the process, even for seemingly positive reasons, may make unanticipated comments that inadvertently share information inappropriately. Such an observation clearly represents a caution for the coordinator in the vignette, who needs to raise the issues brought to her individually in a way that the whole committee can address as a collective, but must do so without unexpectedly putting individuals on the spot. On the other hand, I have found that becoming aware of tensions, discomfort, or conflict in a committee or coalition is helped by gauging the difference between participants’ comments to me in the public forum, compared to their private comments to me either before or after meetings. Similarly, MacNeil (2002, p. 51) describes following up with people after meetings, and using these conversations to read how the deliberative forum is working and to facilitate better deliberative processes.

In discussing a case study of stakeholder conflict, Abma (2000b) notes that conflicts among stakeholders are not necessarily a problem; they can be a resource if they are managed well by the evaluator, and rather than repressing conflict – which could have negative effects – it can be used to surface issues and concerns that ultimately can lead to greater understanding and buy in. She says (2000b):

Dissensus is not necessarily a problem and... it may even become a valuable resource. Taking different meanings seriously into account may initially take more time and energy, but in the long run is certainly worth the effort. When differences are not reduced the chance of resistance and subversive behaviour will be smaller. Another advantage can be found in the increasing amount of creativity. When dysfunctional conflicts are managed adequately, self evident problem definitions are no longer taken for granted. (p. 208)

Abma suggests what amounts to normalizing conflicts and relating them to comparable ones in other settings, keeping a door open to alternative or minority voices, trying to build trust and respect, and generally trying to create conditions in which it is possible to speak across differences, or as Shields (2002) would say, watching out for a community of difference.
On another point, MacNeil’s examples are from larger community settings than I have been working with in my evaluation practice, and the time and amount of preparation required appears daunting. My devolved evaluation contexts involve rapid and frequent turnover of representation from various organizations, as well as potentially significant periods of time between meetings – or at least between opportunities for me to participate in them. Such projects seem less conducive to the form of deliberative forum that MacNeil describes, but the overall premise resonates with the kinds of deliberation that I have tried to promote and build within local community settings.

More generally, the application of ideas about dialogue to evaluation has been inextricably linked to notions of the democratization of evaluation work and the impacts of evaluation (Greene, 2001b; Ryan and DeStefano, 2001; Karlsson, 2001). Implicit in much of this discussion are both teleological and non-teleological conceptions of dialogue. Greene (2001b) views dialogue as a “fundamental value commitment to engagement,” which “can enhance the democratizing potential of evaluation” (p. 181). She further elaborates (2001b):

… dialogue in evaluation contexts refers to engaged, inclusive and respectful interactions among evaluation stakeholders about their respective stances and values, perspectives and experiences, dreams and hopes, and interpretations of gathered data related to the evaluand and its context. The purpose of such evaluative dialogue is to enable stakeholders to more deeply understand and respect, though not necessarily agree with, one another’s perspectives. Such understandings, in turn, can engender more reciprocal, equitable and caring stakeholder relationships in that context, as embodied in an improved, transformed or even revolutionized evaluand. In these ways, dialogic evaluation constitutes an important democratic activity in society. (p. 182)

While the ideals of democracy may represent “the most defensible normative agenda” for evaluation (Greene, 2002, p. 4), they are also not typically the official agenda of those who instigate or contract for evaluation activity. If challenged, few funders would publicly deny the desirability of democratic approaches to evaluation. Yet a democratic approach is
rarely an explicit policy, and indeed evaluation is typically a site of contested knowledge and the imposition of power on the part of those who are able to demand that evaluation be done. Ryan and DeStefano (2001) suggest that dialogue can contribute to civil society “because it assumes stakeholders and the public put aside narrow self-interest and address issues among themselves through respectful, reciprocal conversation” (p. 190). They offer a broad strategy of inclusiveness as the means for accomplishing this, which involves ensuring that stakeholders’ perspectives are represented in evaluation processes (conversations and decision-making), and that the evaluator facilitates this dialogue. Similarly, Greene (2001b) also describes dialogue in evaluation as “inclusive of all legitimate stakeholder perspectives, experiences and value claims” (p. 183). And yet Ryan and DeStefano (2001) note that trying to represent absent stakeholders while also “attempting to maintain a facilitative role can create tensions for the evaluator in a dialogue” (p. 200). Hanberger (2001), citing Gregory (2000) further notes that participation can be based on naïve assumptions about how stakeholders participate in evaluation processes, or arrive at consensus.

Hanberger (2001, also discussed in Greene, 2002) contrasts different democratic models to try and understand the evaluator’s role. Traditional conceptions of evaluation describe the evaluator as an ‘expert,’ which fits elitist or technocratic notions of democracy (for the people). Constrained participatory democracy (by the people) might reflect an advocacy approach, in which the evaluator facilitates a process lead by local interests. While this might be an improvement over the elitist model, it could “promote a fragmented society,” validating a different set of interests, but at the expense of the common good. Hanberger’s suggested alternative is discourse democracy (with the people), in which the evaluator plays a mediating (or even counselling) role, achieving practical deliberation by
facilitating inquiry and joint learning among many stakeholders. Such an approach appears to fit well with the multiple roles and flexible approach demanded of evaluation in a context of devolved programming.

Much of the literature exploring dialogue, deliberation and democratic approaches to evaluation emphasizes the demands on practitioners as encompassing new and challenging roles (facilitator, mediator, advocate, educator) and skills (interpersonal, mediation, negotiation, conflict management, political astuteness). This work is based on a close examination of the nature of power within program delivery and evaluation contexts, and how the evaluator often rides in the nexus of this power – not controlling it, but not flotsam either. Part of what makes these new emphases so important is that they directly address evaluators as a key bridge between the worlds of program practice and the decision-making worlds of funders and policy makers. Such a view of evaluators and evaluation highlights their responsibilities within that context, and reflects on how the default position for many evaluators is to see things from the perspective of their direct clients – and traditionally this has been the funder. Perhaps this is still the case for many or even most evaluators – at least those that primarily focus on large programs that are centrally delivered. But such a view also exaggerates the evaluators’ relative autonomy.

Viewed from the perspective of devolved evaluation, it seems that much of the literature addressing democratic evaluation and deliberation has been written by those who rarely encounter the raw imposition of power – and seems to assume that evaluators can simply pick and choose only projects that meet high-level standards of independence. Putting aside whether evaluators in practice really have such ability to choose, most times when I have been faced with the imposition of power and managerialism it has been more than half
way through a project – typically towards the end. By that point in the process, simply backing out of the project is not a useful or even feasible option for either the evaluator or most stakeholders. Or as Cervero and Wilson (2006) note, “Once engaged … evaluators become stakeholders in the program and must carry out their work in the matrix of power relations that structure the social and organizational context” (p. 234).

Perhaps it is just a matter of upfront negotiation about the parameters of the project and the roles of the evaluator and other stakeholders, or simply choosing our projects carefully? In exploring the role of evaluation in democracy, Henry (2001) suggests:

Rather than carrying out studies assigned by a central authority with results controlled by them, we have a range of constituencies and great latitude about how we choose to do the work of evaluation. Because evaluation is a relatively young field populated by clever and thoughtful people, options for evaluations abound and there is little consensus to constrain our practice. (p. 427)

And yet there are constraints on all evaluation practice. Such constraints are particularly implicated in contexts that provide some evaluators with a great deal of discretion and provide others with little. Clients – either funders or those in local communities – are interested in maximizing their ability to control what takes place in the evaluation. In devolved evaluation settings the opportunities for gaining a wide audience that can safeguard the evaluation and evaluator from managerialism or from narrow restrictions can in some ways be more difficult to obtain. The prospects for broadly based inclusion are challenged by the time frame and scope of program implementation, as well as the number of potential participants in such a process, and so I find that I must marshal a range of supports to address issues of control over the evaluation process; anticipating that such challenges are part of the terrain allows me to develop a variety of viable processes for preventing the most flagrant impositions. Such supports can include a proactive approach to widening the decision-making sphere – as broadly based a community coalition or steering group as possible;
backup in the form of colleagues or evaluation partners who are clearly part of an advisory
team, if not directly involved in implementation; and symbolic and instrumental supports in
the form of membership and reference to codes of practice of the Canadian Evaluation
Society (CES), and the American Evaluation Association (AEA).

Part of my shift to working with nonprofit organizations has reflected my frustration
with demands of corporate clients who find many ways to restrict my independence.
Working in a more openly mediating situation can allow me to surface those tensions in ways
that can mitigate the power of the funder, partially because of the more public nature of the
actions of all stakeholders involved. It can allow me, and also my community level clients to
negotiate with more leverage. While managerialism can also occur with local community
clients, it tends to be more straightforward to cope with, because there is virtually always a
funder or other powerful stakeholder balancing such efforts, and providing the necessity for
openness concerning the products, if not the process of the evaluation work.

Another way that I have negotiated leverage in community level work is through
being associated with larger organizations – university research centres and interest groups,
larger associate consulting firms, or larger groups of independent consultants who are
capable of providing moral and intellectual support, as well as the ‘appearance’ of being part
of a larger collective. Such groups can give my work legitimacy among local stakeholders
and funders, and obviate some of the tensions of being an independent evaluator potentially
more susceptible to managerialism.

Beyond the direct imposition of control on the evaluation process just described, there
are many forms of indirect control manifest through the definition of problems, solutions,
organizations and communities. The assumptions funders make about how communities
experience problems, and how they can and should cope with them have profound implications for the work of evaluators. In particular, the conceptions of community – their boundaries, cohesiveness, capacities, and needs – are reflected in how programs are developed and implemented. Recent work by Shields (2002) and Furman and Starratt (2002) examining ‘communities of difference’ challenges prevailing ideas about the nature of community – ideas that underlie and implicitly inform traditional approaches to evaluation. Efforts to open up evaluation decision-making beyond the narrow scope of funders introduces a concomitant need to address the diversity of voices that then may enter into discourse about the worth of programs. Shields and Selzer (1997) propose:

… that a more adequate concept of community will of necessity be more complex, more dynamic, and more public – grounded in notions of dialogue and processual strivings. We have recognized that dialogue among individuals who respect each other deeply and who wrestle with the meaning of culture, power, empowerment, and community may be as complex as the context itself. (p. 435).

Shields (2002) and Furman and Starratt (2002) further problematize the concept of community in several ways. The first and perhaps most crucial way is that most conceptualizations of community emphasize the commonalities among community members rather than appreciating the differences within every community and group. The following vignette builds on this basic idea to explore some of the implications for programs and for evaluation of rethinking the nature of community.

**VIGNETTE 6.4.2 (GRASSROOTS & NONPROFIT) – COMMUNITIES OF DIFFERENCE**

We are nearing the end of the eight-community demonstration project, and are taking advantage of the opportunity represented by a local conference to bring together representatives from each community to discuss the projects in a panel forum. The conference theme crosscuts the various projects quite well, and the representatives from each
community – my evaluation team members – are excited about the chance to share what they have learned. I will be facilitating the panel, and will be sharing the stage with ten individuals representing the eight project teams. The format will entail each individual making a brief statement describing key goals, activities and learning from their projects, followed by an extended question and answer discussion with the audience.

Although representatives from the communities have been brought together in a central location on two prior occasions, due to recent turnover of project representatives three of the ten people have not yet met any of the other community representatives except through conference calls. Indeed, I have not yet met these individuals in person, and I am somewhat apprehensive about how they will represent their projects and communities.

As the participants begin describing their projects, I am amazed at the group cohesiveness displayed by the ten panellists. Their enthusiasm and camaraderie is clearly evident to the audience, and have obviously been extended to include the new individuals who were introduced just moments before the panel session began. The representatives do not display a uniform understanding of the project activities, rationale for what has been undertaken, or implications for practice. Indeed they express disagreement on occasion, and at times their discussion takes on the character of a long-term and spirited debate – one that they appear to have revisited on numerous previous occasions, although it does not seem to be staged, and feels quite authentic.

They support one another in describing and remembering key events and episodes from one another’s projects, and the longer-term representatives suggest and tell stories heard about other projects on behalf of the newer members who had not yet heard them. They spur one another on to describe important details and help weave a collective whole from very
disparate stories and experiences – “Talk about the community workshop last September,” or “Our experience was like this, but in April’s group, they had a different experience … do you want to talk about that, April?”

Two of the groups have more than one representative from the project, and it becomes clear that the individuals do not necessarily agree about everything that happened in their projects, or how the situations should be interpreted. The audience members get into the spirit of the discussion, and ask questions aimed at illuminating the different shadings of understanding and meaning within and between projects. One of the most powerful discussions involves contrasting two communities that are geographically quite close, similar in size, population and occupational distribution, and that have implemented several very similar program activities. The contrasting descriptions of the ways the project activities were implemented, the extremely different experiences of ‘success’ and ‘failure’ that the community representatives describe, and the very dissimilar next steps outlined are a revelation to panel members and the audience. The two communities are usually considered so similar and uniform in character that they are often described interchangeably.

**Reflection 6.4.2**

Furman and Starratt (2002) challenge the ‘sameness assumption’ about communities by noting that peoples’ experiences of community are often temporary, tentative, transitory and pragmatic. They emphasize community as constructed or invented ‘on-the-spot’ rather than relatively stable and ‘found.’ They further note that traditional conceptions of community can lead to balkanizing solutions to problems – when we end up putting people into groups based on their similarities, that’s who they end up talking to; they don’t hear about alternative scenarios. In my experience doing evaluation among local communities,
governments and program funders tend to disregard the diversity of perspectives within communities, as well as the diversity within program communities, programs, and service delivery organizations. Perhaps even more problematic is that government tends to ignore the diversity among communities as well, treating one remote community much like others, regardless of the differing historical and professional experiences of community members.

In the vignette, the panel and audience of the conference session are challenged by the dissimilar experiences of two seemingly nearly identical communities. The opportunity to share stories about their experiences in this way allowed the participants to move beyond simple stereotypes to understand the unique attributes of the individual communities, and how these were relevant to program implementation and evaluation. This is particularly critical given the common practice of having evaluation efforts in even the most unique and innovative projects focus on ‘best practices’ and ‘lessons learned’ – with the goal and assumption that the learning can and should be applied in other communities.

For Shields (2002), a ‘community of difference’ emphasizes not the high level of commonality usually associated with 'community', but rather focuses on

… explicit, new, and agreed upon understandings…. the commonalities are values of inclusivity, respect, and a desire to understand diverse perspectives; the norms are commitment to reflection, critique and dialogue. (p. 213)

Shields (2002) and Furman and Starratt (2002) use the ‘community of difference’ reframe of our understanding of community to reflect on a variety of implications for the study of leadership in community contexts. This work identifies a critical role that evaluators play by virtue of their social location as a mediator, facilitator and participant in the worlds of both practice and policy; the worlds of implementation and funder.

Shields (2002) argues that such values as developing respect, eliminating power inequities, embracing diverse perspectives, and establishing high expectations are central to
leadership for communities of difference, and represent the new ‘common link’ for communities, suggesting a way of turning what might have been seen as challenges or deficits into strengths. Certainly the local communities in the vignette benefited from having the opportunity to contrast their settings, problems, strengths and activities with others in their own and other projects. They gained valuable ideas, while also having a ready yardstick to gauge such issues as their own responses and plans for action.

During the project described in the vignette we allocated a period of time during our communal gatherings for sharing stories and ideas about the projects. At the first of two meetings, it quickly became clear that the brief time allotted was inadequate, and that the representatives wanted and needed more space in which to work together, and get to know and support one another in understanding their own projects and the funder’s expectations.

Providing leadership in evaluation using a ‘community of difference’ approach involves being able to recognize the context and the differences, and supporting and facilitating those differences as legitimate and appropriate. It means developing and depending on community members to provide critical leadership, as otherwise participants will have little buy in for the process beyond a minimally required emphasis that might meet the needs of funders. It means working to identify and reinforce shared values and goals within the community. It means being sensitive to marginalized stakeholders and introducing inclusive measures to invite them to a seat at the table. It means being able to speak to a broader constituency – linking the work in the local community to what is happening in other communities, or organizations within the community.

Another way that Furman and Starratt (2002) problematize our conceptualization of community is by reflecting on how communities are typically seen as isolated or at least
treated as such. Viewing communities as isolated contributes to seeing them as unequal or disadvantaged, and this is exaggerated for remote rural areas in contrast to larger urban ones. It translates into a dependence model in which direction and control must come from the centre, as the centre looks to the regions and can't conceive of local resources capable of dealing with issues and problems.

My recent work with local communities has reinforced for me the breadth of local resources that most communities are able to command when given the opportunity and the need. In this context I have found that there can be a strong role for an evaluator who contributes to local community members' perceptions of their own capabilities, but also supports this with active resources that fill local gaps. And part of this role is to facilitate communication and support network development among communities, as well as within them. With the eight-community demonstration project such an approach was implicit in the model, but it applies to any community context, as making those links to near or distant communities builds that connection and raises the bar as communities raise standards and expectations, and selectively appropriate and adapt aspects of ‘what works’ elsewhere.

Swindell (2000) argues that devolution is potentially a false answer to questions of diversity – by devolving until you reach homogeneous groups, or at least what would appear to be homogeneous groups, you remove many of the possibilities and opportunities for dialogue that strengthen those programs and services. Swindell notes that one rationale for devolution is to increase the sense of democracy through expanded citizen participation. Yet he is concerned that this is an untested assumption, and that devolution does not necessarily move the locus of decision-making closer to the people, and even if it did, that this could overcome public apathy or distrust of government. The community of difference model
would further add to Swindell’s argument by noting that even when we think we have
devolved to the level of homogeneous groups, significant diversity likely remains.

One of the common elements of approaches aimed at promoting democratic
evaluation, or looking at the democratic and deliberative impacts of evaluation practice is to
focus on the ‘common good.’ Helping people define ‘the common good’ around specific
issues can mean raising the bar, and linking to extra-community goals and issues. Yet most
writers do not adequately unpack the notion of the ‘common good.’

I believe that assuming that it is always possible to identify or at least come to
agreement on what is the common good leaves it open to being co-opted or appropriated,
which is problematic for a number of reasons. First, when real and political issues are being
contested, addressing them can involve long-term and negotiated efforts to even reach
agreement on defining issues, let alone resolving them or reaching agreement on the common
good. Second, because of the transient, fluid, and multiply overlapping constituencies of
which communities are now made (Furman & Starratt, 2002), there are always going to be
ongoing unresolved issues of having all the voices at the table, and ensuring that all the
relevant voices are heard. Which leads to a third point, that the outside world impinges with
its competing definitions of the common good, and processes for achieving it. It is not
uncommon for those working within a local community context to experience frustration at
their position as a go-between for local, provincial and federal concerns, which are not
necessarily about the same issues, and usually involve significant differentials in power.

The broader community in which the program or geographic community resides has
interests, processes and values that impact on the community, over which local communities
may have little control. The question for local evaluators who are trying to implement
democratic community approaches becomes “how is it sustainable if you can't deliver the goods?” How can even the best local processes contend with the imposition of values and ideas on communities from outside? It is possible to envision a rich and rewarding, but intense and hotly contested dialogue about wide-ranging issues, in the end usurped by decisions made elsewhere, addressing issues that have not been locally validated, using processes with which the democratic community was unaware and to which it was not a party. Herein lies a powerful and important role for evaluators.

Implicit in the discourse on dialogue and democratic evaluation is a sense of the role of the evaluator as somehow, tacitly and skilfully taking responsibility for evoking a respectful, inclusive, equitable, balanced, fair, tolerant, and open environment and dialogue; perhaps in the face of resistance, usually in the face of competing (and sometimes incommensurable) interests, and always within a context in which the evaluator’s role is contingent, ambiguous, transient and emergent. Missing from most of these discussions is how the evaluator does this, learns how to do this, and obtains the right, the permission and the opportunity to undertake such responsibilities. This has been a key task that I have chosen to take on with my practice over the past decade, exploring and expanding this nexus, and through this dissertation, trying to improve on what I do. It has involved exploring the nature of leadership in evaluation practice, working to understand the realities and the potential within my work as an evaluator, and their implications for broader evaluation practice.

Summary

Chapter Six examines evaluation practice in the context of devolution. I reflect on four broad and related themes encompassing trends within evaluation theory and practice – theory-based evaluation, inclusiveness of practice, the diversity of non-technical roles
undertaken by evaluators, and efforts to introduce dialogue and a deliberative democratic orientation to evaluation work. In turn these trends are at least in part both reflective of and constitutive of the changes in context examined in this dissertation. The devolution of program and service delivery to nonprofit and grassroots organizations has been mirrored in the devolution of evaluation responsibilities to these organizations. In my practice I have wrestled with the implications of this potentially exponential growth in the number and diversity of interests represented by those who find evaluation to be relevant to their work and responsibilities. Recently I have seen evidence that the discipline is addressing some of these implications, although the discussion typically appears submerged within other long-standing and current debates.

My emphasis has been on understanding the potential and actual roles that evaluators can and do play in working with nonprofit and grassroots organizations. Chapter Five examines devolved evaluation through the lens of the impacts of devolved programming on the context of programs and services, and as such focuses on such activities as negotiating accountability requirements, providing continuity for organizations amidst turnover, burnout and stress, observing and documenting mandate drift and its implications, mitigating some of the repercussions of competition among and within program stakeholders, and making sense of the complexity of contexts represented by multiple community funders, multiple funders, and the myriad permutations of program design and delivery that ensue from this.

In Chapter Six the frame and my task are reversed – starting with broad trends in evaluation, I examine the context of devolved programming as special cases in which these trends are apparent, but also challenged in specific ways that potentially offer insights into broader evaluation practice. And again, with the emphasis on my work as an evaluation
practitioner, I examine recent trends toward theory-based evaluation, the process and potential for building participatory capacity in nonprofit and grassroots organizations, the diversity and central bridging roles that evaluators can play in such settings, and the challenges of promoting and enabling dialogue and deliberation among, within and between communities, stakeholders, funders and the academy. Central to all of these endeavours is considering the nature of the negotiation process, and the many ways that what evaluators do is not simply evaluation, but participation in broader community and organizational development, with implications extending far beyond immediate evaluation concerns. Chapter Seven continues this reflection on the place of this study within the broader discourses about both evaluation and the devolution or programs and services.
CHAPTER SEVEN:
OFF THE SIDES OF THEIR DESKS

In the persistent difficulty evaluators face in accomplishing (this) lies a latent source of unfairness and injustice built into the fabric of program evaluation, for it tends to favour the voice of those few for whom programs are useful instruments to advance their careers and their economic power. For the majority of people implicated by or involved in a program, the concept of ‘program’ is barely understood and may even be irrelevant to their lives.

Saville Kushner, 2000, p. 10.

The devolution of programs and services to local communities has profoundly affected my evaluation practice, and changed the locus for evaluation activities. Evaluation is being demanded of nonprofit agencies delivering programs in local communities. As such, the scale of the research tends to be smaller in scope, more often done by administrators or program deliverers, and perhaps less often by ‘professional’ evaluators, with the exception of people like myself who are willing and able to take on numerous small contracts with multiple organizations. Even so, in such situations I often do not do the evaluation for the client so much as facilitate it; using a participatory model, I do parts of the evaluation, and coordinate and assist others in the organization as well as volunteers who take on tasks that can encompass virtually any aspect of the data collection and planning processes.

With community-level nonprofit and grassroots organizations as clients, I often deal with new or pilot projects, which may be funded for between one and three years with an expectation that the organization evaluate program activities. Few large-scale programs are evaluated in the first year or two of implementation. When they are, the focus is overwhelmingly formative, to give feedback about how to adjust or fine-tune the program. Devolved programming is usually not given this grace period in which community practitioners can work out challenges; they are expected to demonstrate immediate program effectiveness. As such, part of the responsibility for evaluation in a devolved setting is to
make sure that projects maximize opportunities to prove themselves, undertaking as rapid a
shakedown of implementation issues as possible.

In such contexts the evaluator of necessity takes on many roles, as do most people
working in these agencies and community level organizations. Such roles fundamentally
change the nature of the relationship between evaluator and all stakeholders; roles are fluid
and emergent, transient and context-dependent, and inevitably require leadership in action
and leadership in simply recognizing the possibilities and needs for action. Indeed, the cases
and vignettes presented in this dissertation build an image of the evaluator as someone who
crosses frames, and acts both within local contexts and on those contexts to change the nature
of the relationships among players. I contend that evaluators in this sense are more than
technical experts or researchers; they are players in a complex array of stages and contexts.

In this way I observe how recent changes to the context of evaluation practice
enhance the role of evaluators as players in the political discussions about the desirability and
need for programs. Such roles already exist in most evaluation practice, but the devolved
context of program delivery and evaluation extends these roles, and make them more
transparent. The case studies and vignettes describe a context in which stakeholders are
actively taking part in and appropriating evaluation activities. These activities are
fundamentally shifting power that has traditionally rested with nonprofit organizations out of
communities and to funders. This reciprocal flow of power is negotiated and appropriated in
a long-term dialogue among political stakeholders, administrators, deliverers, and program
clients, and implicates the changing roles of evaluators within the nexus of this negotiation.
Before further taking up this intermediary and negotiation role of evaluators, I will briefly
summarize and return to the study’s original reflective questions.
Revisiting the Reflective Questions

My reflection on practice in this dissertation began with the following questions:

1. How and why has my evaluation practice changed over the past twenty-five years?

I began my evaluation practice working as a junior consultant on evaluation studies of national programs delivered by employees of the federal government. I moved into government and as an internal evaluator examined provincial level programs designed and delivered by government employees. During this period, I saw the beginning of a move toward devolved program delivery, and became involved in several efforts to promote evaluation activities within nonprofit sector organizations. Returning to the consulting world, my practice initially involved evaluating national and provincial programs delivered by public servants and by nonprofit organizations. Over the past dozen years as a consultant my work with grassroots and small nonprofit organizations grew to become the focal point of my practice, although in partnership with several larger associate consulting firms I continued to undertake some evaluation work on federal and provincial level programs. However, even these programs are typically delivered by nonprofit organizations, or the delivery is of a policy, strategy or funding program rather than a service provided directly to clients. Some of my evaluation related activities have remained similar – data collection, report writing – but in most cases the work has expanded to encompass a diverse variety of additional tasks, roles and responsibilities, and has come to encompass strikingly different relationships with an expanded set of stakeholders and clients. In particular, most evaluation projects that I have taken on in the past ten years have involved working directly with deliverers in collaborative and participatory ventures, in which my specific role sets have been fluid, customized to the needs and capabilities of my clients, and often ambiguous. This ambiguity of role has been challenging in part because I have not been able to find many examples of precedents and
role models. Indeed, the traditional evaluation literature has raised numerous cautions against undertaking many such roles and activities, even as the practice context has made it clear that new ways of operating have been necessary. I undertook study in the Ed.D. Program to help me better understand the nature of the changes that have been so obvious in my practice.

2. How does the devolution of programs and services affect the practice of evaluation?

The first programs and services devolved to the nonprofit sector were evaluated using traditional approaches, and often involved consultants or internal government evaluators undertaking extensive field work with community delivery agencies. More recently the evaluation responsibility itself has been devolved to nonprofit organizations, giving them the choice of evaluating agency level programming with internal resources, hiring someone to help with the evaluation, or recruiting appropriate volunteers to take on the task. The accountability focus has been paramount in such settings, whether or not it has been the primary or manifest intention of the funder, meaning that devolved evaluation has involved a strong emphasis on self-regulation with respect to the delegated authority. This devolved evaluation context has provided such challenges as competition among deliverers, inadequate resources to undertake evaluation tasks, and a complexity of roles, relationships and demands of a diverse stakeholder network. In short, programs are becoming more challenging to evaluate even as the capacity to do so is compromised and the requirement enhanced.

3. What are the problems and challenges of conducting evaluation in a context of devolved programs and services?

Beyond the local capacity of organizations to undertake evaluation, resource capacity represents a significant barrier to organizations hoping to expand their capability. I have addressed this in part through developing collaborative and participatory approaches to tasks, doing only those parts of the evaluation that cannot be done by someone within the
organization. The increasing complexity of devolved delivery settings further challenges the
capacity of organizations and of external evaluators endeavouring to understand the
programming environment. The tasks require more than just being a fast study, but
sensitivity to the potentially charged political environment, and a willingness to live with the
ambiguity inherent in such complex and constantly shifting terrain. The fluid delivery
environment can mean that examining the same program two years apart in time can mean
dealing with different delivery staff, agency and perhaps funder. In such a setting the
evaluator can be a key link to a program’s history, and have a much broader responsibility to
the community and programming environment than simply to one individual evaluation.

4. How does evaluation affect devolved programs and services?

The increased requirement to undertake evaluation challenges the capacity of most
organizations. But those working in such settings often do not resent the imposition, despite
its apparent emphasis on accountability, and the enforced self-regulation. Most of the people
I work with in grassroots and nonprofit organizations welcome the opportunity that
evaluation represents – to showcase their efforts, demonstrate the value of what they are
trying to do, and to learn about how their efforts can be improved or contribute to learning
relevant to other organizations or communities.

Increasingly, as I have learned how to build linkages and facilitate connection among
stakeholders, I have come to see the potential for evaluation to become a force of community
development rather than of regulation, assisting with the reconnection of programs with other
communities, ideas, resources and processes. And yet at the moment, the most common
impact of devolved evaluation is that community efforts are isolated and fractured, and what
was a growing transparency of programs has been reversed. Even so, the evaluator can be in
a position to address such divisions, and to establish linkages in the face of competition,
turnover within agencies, and a haphazard and complex funding environment. And yet, depending on the skills and good will of evaluators to take on such roles in the face of reduced resources, unclear and potentially antagonistic mandates, and a disciplinary focus on technical mastery over adaptive ingenuity and political astuteness limits this potential benefit.

5. What are the implications of devolving evaluation to nonprofit organizations for clients, programs, organizations, communities, evaluators and the broader society?

For the grassroots and nonprofit organizations delivering devolved programs and services, the evaluation requirement has diverse and contradictory implications. Most know the obvious benefits of evaluation work, yet the process is fraught with potential dangers. As evaluation becomes part of the decision-making process employed in establishing and justifying funding decisions, the need to do evaluation and do it well increases in importance. Yet complex and competitive contexts make it more difficult to do evaluation well. Further, pressure to standardize program goals, success criteria and evaluation approaches has the potential to confiscate from community agencies the opportunity to demonstrate the unique benefits and value that their organizations have within their community settings.

For the broader society, this can mean lost opportunities for program innovation and for deliberation about the value of programs and services, and reduced transparency of government expenditures as individual organizations focus on their own work outside of the broader societal context in which they are funded and situated. Yet when evaluators bring in new tools such as participatory approaches, focus on promoting dialogue and deliberation about programs among stakeholders, and actively embrace diverse roles in mediating among community players, funders and other interested parties, evaluation has the potential to provide leadership in accommodating the demands of the new delivery models. The next section explores some of this potential in more detail.
Implications for Practice

The starting point for the reflection on practice undertaken in this dissertation is the transformed context of devolved program and service delivery. This does not appear to be a short-term shift – returning to a delivery model employing primarily public servants would seem to represent an even larger shift than has taken place over the past two decades. However, the devolution of evaluation responsibilities is an unfinished trend. Indeed, while I argue that the locus of much evaluation practice has been actively changing, devolved evaluation is by no means a consistent or complete process. Traditional evaluations are still being conducted – even if program delivery contexts may be more complex and varied than in the past. If anything, the devolution of evaluation appears to represent an expansion of how evaluation work is done as much as a shift – more evaluation is being done, not less.

The recent accumulation of critical literature on the impacts of devolution on nonprofit organizations reveals issues relevant to all evaluators. This literature suggests that devolution has implications for who contracts evaluation, who conducts it, the range of stakeholders relevant to evaluation work, and the uses made of evaluation findings. As documented in the vignettes and cases in this study, devolution also has implications for how evaluation is conducted, and the nature of the activities considered within the purview of evaluators. My concern in particular is that devolution has contributed to the creation of gaps in our processes for understanding and reflecting on the successes, struggles and impacts of programs and services. I see two broad ways that these gaps play out. One is that evaluation as it is typically conceived and conducted does not and perhaps cannot examine the impacts of the process of devolution itself. To do so would require a process of meta-evaluation encompassing multiple program delivery settings over a period of time, and span diverse
evaluation contexts and studies both within and across communities. The other is that the conduct of evaluation is itself a context for evaluation practice – the impacts of evaluation on programs, organizations, communities, and the broader devolved programming context needs to be critically examined as a site of contested knowledge construction and use.

For example, Alexander et al. (1999, p. 464) argue that local “organizations have a kind of knowledge authority based on their familiarity with the community and the community's trust in them; they are of the community.” When the state devolves responsibility for delivery of social programs to local communities through these organizations it is trying to tap into that trust and that local connection. Alexander et al. (1999) and others note that some of the unanticipated consequences of doing so include a diminishing of the trust and community connection that has been a major part of the rationale for devolution. For evaluators examining these programs, with either the state or the local community as ‘client’ for the work, the devolution itself and any unintended consequences of devolution of necessity become part of the scope for the evaluative gaze. To maintain a narrow evaluation scope on a program’s implementation divorced of its embeddedness within this devolution context short-sightedly neglects powerful influences on the program’s putative success or failure. Changes to context are at least as relevant to the program’s implementation and success as are its other salient delivery characteristics.

Eikenberry and Kluver (2004) provide another example reflecting the implications of evaluation within the program context. They suggest that public-sector administrators can help play a role in mitigating the impacts of devolution within the sector. They argue that administrators need to understand the role of the nonprofit sector in a democracy, the impact of devolution, and the need to support interorganizational coordination and collaboration:
... public managers can play a significant role in enhancing nonprofit organizations’ ability to add to civil society by supporting collaborative and cooperative efforts among nonprofit organizations rather than promoting competition. Collaboration and cooperation can be especially powerful in retaining the essential characteristics of the sector. When nonprofit organizations work together, they empower themselves to retain their unique values, focus on service and advocacy, and maintain civic involvement. (138)

As part of the broader administrative apparatus – *de facto* if not *de jure* – evaluators can contribute by recognizing the impacts of competition on communities, helping to re-establish and maintain connection and collaboration among community players, and fostering awareness of the contextual embeddedness of local programming. Further, they can examine their own role as part of the accountability apparatus that creates and reinforces this competition among organizations, and is thereby implicated in the success of programs, agencies, and the broader delivery environment. Evaluators can also recognize that employing participatory approaches to involve deliverers in evaluation further implicates them in their own self-regulation.

Ebrahim (2002) provides another example. His study of struggles over information between international nonprofit organizations and their funders portrays the relations as a form of “resource exchange: the exchange of information for funds” (p. 107). As a site of contested information development, negotiation and resistance, Ebrahim argues that the exchange serves to structure the relations themselves, and in effect reproduces and entrenches those relationships and the information systems on an ongoing basis, furthering and entrenching the tensions between organizations and funders. He argues that easing back on the type and amount of information demands could result in improved efficacy of the organizations, higher levels of innovation, and responsiveness to community needs. Such a perspective mirrors the kinds of impact of demands for accountability and evaluation described in this study.
While I believe that we need to examine the scope and subject of evaluation work in a context of devolution, I also perceive it is an opportune time in which to do so. Phillips (2001) has argued that we are entering a time of reform of state-voluntary sector relations, and that there have been some signs of movement from more hierarchical to collaborative forms of governance. If this is so, Phillips suggests that this shift will require a number of institutional reforms that have significant potential implications for evaluation practice. The reforms that Phillips suggests include: 1) enabling the nonprofit sector to fulfill its distinctive roles, 2) engaging sector organizations in genuine and meaningful dialogue, 3) elasticizing the means and practices of accountability, which would introduce a measure of flexibility and responsiveness into the accountability relationship, and 4) embodying horizontal management policy across departments and across governments using new governing machinery, and not merely through *ad hoc* processes (p. 182).

At present, such changes appear to be inconsistent, if the pattern is actually real. Thus far, as accountability mechanisms of hierarchical governance have been extended to the nonprofit sector, they appear to have increased centralized control over programming. Indeed, Suda (2006) notes that devolution involves a seeming contradiction – that government’s role decreases with the replacement of public provision of programs and services – characterized as ‘shadow state,’ ‘government by proxy,’ or ‘hollow state,’ – and yet in terms of regulation and control, the role and influence of central governments seems to be stronger (p. 445). While accountability requirements do not constitute the sole rationale for evaluation, they do represent perhaps its primary driving force within the nonprofit sector. As nonprofit organizations respond to these increasing demands, evaluators need to reposition their work within this new matrix.
The context of devolution sets in bold relief some key tensions and pressures within broader evaluation practice. Contrasting the specific aspects of context that seem to be unique to devolved programs outlined in Chapter Five with the broader trends in evaluation discussed in Chapter Six suggests several implications relevant to all evaluators, some of which are theoretical, and many of which are quite practical in import.

One way of thinking about the devolution context and evaluation is to ask whether some of the features of how devolution has been experienced by nonprofit and grassroots organizations are inevitable. Certainly not all such organizations experience all of these features, or at least not all at once. The accountability pressures seem to be the most universal and ubiquitous – they pervade virtually every aspect of what such organizations are doing, even when they are hidden within other activities and requirements. Whether it is an accurate reflection of the intent, in practice most nonprofit and grassroots organizations equate evaluation with accountability, and usually within the frame of accountability as answering to the funder as the delegated representative, and through this, accountability to the public for the appropriate expenditure of scarce resources. The case studies and vignettes provide examples of how efforts to clarify program and project activities by using logic models can effectively ossify program elements and reduce their ability and tendency to adapt activities to the needs of clients, making programs less effective and relevant. That stability itself becomes a part of the accountability requirement, whether it is appropriate or possibly damaging to the ability of the agency to meet client and community needs.

Similarly, the drive for external accountability has enlisted project and program workers in various methods of self-regulation – entering program data into databases, doing client follow-up, examining agency activities through participatory approaches to evaluation
which can engender active resistance, implicate evaluators as part of the ‘ruling apparatus’ and a culture of surveillance, and thereby make all evaluation work that much more challenging to accomplish. It is not difficult to understand resistance to evaluation when the activity is so closely tied to very real decision-making processes about funding, when it is something done to agencies and programs, or done by staff to themselves, and is virtually always a top-down requirement rather than growing out of a local desire to understand and learn about ‘what works.’ This is further exacerbated when the products of that evaluation—the reports, the data, the outcome of so much effort—appear to disappear into a black hole from which no positive consequences seem to emerge. They rarely seem to contribute to public discourse about the worth of programs or new approaches, they can contribute in a negative way to decisions about the withdrawal of funding, but perhaps the most disastrous outcome is when reports disappear into that black hole and beliefs become reinforced that no one really reads, cares about, or has any intention of using the reports or findings.

The question that comes to mind for me as I examine this minefield is whether it would be possible to distinguish evaluation goals from accountability goals? As a tool of accountability for funders, evaluation is blunt, often inappropriate, and fraught with unintended consequences that can generate outcomes the opposite of what is intended (see Sieber, 1981; Smith, 1995; Tenner, 1996). At its most egregious potential in the performance measurement movement, this accountability expectation conflates means with ends, and the process itself can have impacts on programs that make them less effective, change their direction (mandate shift), challenge the capacities of organizations, and bring on active resistance as well as avoidance, misdirection and sabotage of legitimate and possibly helpful
evaluation activities. Distinguishing evaluation from accountability goals might be challenging, but it is worth considering how that might be accomplished.

Another facet of devolution that could be mitigated is the capacity challenge. The short-term funding cycles with high accountability and outcome expectations reduce innovation and flexibility, while increasing the competition pressures from other agencies. Scriven (1991), in discussing the rationale for goal-free evaluation, addresses the competitive pressures of organizations, noting “a gallant try at Everest will be perceived more favorably than successful mounting of molehills” (p. 57). In this way the quest and competition for resources helps to obscure the goals of programs – reinforcing the gap between ‘official’ and ‘operational’ goals (Chen, 1990, p.174), and also having an unintended consequence of encouraging every organization and program to make what it does look different in some way from what other organizations do. Thus efforts to compare across projects or develop a more comprehensive vision of what is being delivered are made more difficult.

The complexity introduced by having multiple and joint funders for programs and agencies further muddies these waters by introducing competing and sometimes contradictory expectations about evaluation, accountability, reporting and timelines that reinforce capacity challenges for organizations, which affects what they are able to provide and their willingness and ability to adapt and remain flexible to client and community needs. The lack of coordination between diverse funders, particularly with respect to evaluation and reporting expectations, contributes to these capacity challenges, but better coordination could bring about its own challenges – new standards enforced from outside communities, improved timelines that could remain out of touch with the realities of local delivery, and so on. Providing local deliverers of programs and services with better opportunities to provide
input and coordinate directly on such matters with all funders and other deliverers could help, but such participation itself represents a capacity demand that needs to be acknowledged and supported. Similarly, the administrative demands of evaluation and reporting clearly represent a fundamental part of the delivery landscape, and rethinking and redesigning the eligibility of such activities for core funding is part of the needed capacity transformation.

While there may be some ways that the context of programming and evaluation in nonprofit and grassroots organizations might be changed, the presence of evaluators and what they do in such settings can make a difference. Beyond the obvious knowledge and expertise capacity needs within small organizations, evaluators can and do provide more than technical skills. By using participatory approaches – in whatever form the organization is able to take part – evaluation contributes to capacity building through education and action. By working with diverse stakeholders, evaluators can identify and address inequalities, represent, mediate, facilitate, negotiate and otherwise provide linkages between different interests and concerns. By striving for rigour in research and the conceptualization of programs and their impacts, evaluation builds and supports improved levels of programming and helps define standards of practice that benefit clients and organizations. By facilitating meaningful discussion and deliberation among organizations and between those representing program governance and delivery, evaluators can build networks that connect and reconnect practice with theory and policy, and address the fragmentation that arises from devolution.

In practice, I see these roles and activities not simply as hypothetical goals to which I think evaluators should quest, but actual de facto activities that many evaluators now need to do in order to do their jobs well. But often the action of taking on these roles is implicit – it may even be concealed. The roles are not well defined; they are demanded of the context, but
Although I have been profoundly impressed by Schwandt’s (2003) vision of the ‘rough ground’ of evaluation practice, and the need for evaluators to cope with paradox, ambiguity, constraints, contingencies, contradictions and contentious issues (pp. 355, 361), I am concerned that this vision depends so heavily upon relying on evaluators to do the right thing – to be aware of the impacts of their actions as part of the context that creates constraints for organizations. Finding a space in which evaluators can legitimately undertake social critique means changing the contexts in which evaluation work is done, and not just evaluators or what they know. For this reason I have come to reflect on evaluation as not just a profession, but a societal activity, one that is capable of addressing our need to improve our social interventions, but also address broader concerns of social justice, inequality and
differential power. Schwandt discusses how evaluators can legitimately examine the norms, values, and beliefs informing practice – and how they are typically considered off-limits except as the subject of empirical study (p. 502). While some view the role of being a social critic to be appropriate for evaluators, many express concerns over the ‘objectivity’ of their work in doing so. But in viewing the diversity of roles that evaluators undertake in nonprofit and grassroots evaluation, I believe that in many ways they are uniquely situated to do so – they are in a situation, but not of it; they straddle many contexts at once, and must rise above each while being effective in working with many stakeholders on their own terms. Evaluators in this way have access to information, have opportunities to see and participate in relationships, and actively represent people to one another. It is a setting in which evaluators have power, but one that demands a great deal of tact, sensitivity, and a breadth of perspective that encompasses diverse interests, and an orientation that goes beyond the individual study, program, organization and context. In very practical terms, to view evaluation work in this way means acknowledging the potential power and potential for harm that can ensue when the work is done either well or poorly. It represents a potentially huge moral landscape – one that could demand a wholesale reconsideration of how we teach, certify and regulate the activities of evaluators. In the short-term it may be enough to recognize that such needs and such roles already do exist, and to explore the implications.

One positive implication that I find most intriguing is to think about meta-evaluation. As I have noted at many points throughout the dissertation, I find that evaluation in a setting of devolution demands that I continually be aware of contexts and programs and stakeholders outside the obvious and local frame. I am continually linking people: communities (doing similar work, interested in the same or related issues, programs, or theories), academia (as
resources, but also as interested parties to the social learning taking place), funders (always searching for the perfect program, idea, efficiency, ‘lesson learned,’ ‘best practice,’ or the holy grail of social programming – ‘sustainability’) and individuals (program stakeholders, clients, other agencies or programs, the un-represented, the dispossessed). My efforts to use participatory methods involve long-term capacity building that may not benefit the current project. The program I am evaluating may make little sense outside of its situation within an agency, a community, or a funding cycle. My efforts in one agency may provide me with an opportunity (or an obligation) to represent the corporate memory for a program, agency or set of community-wide activities, when I can provide the sole link to previously delivered activities or possibly long-lost organizations or programs.

The problem I perceive here is that so much of what is learned in the short term is lost to the long term. My reports are often considered confidential, so even if the agency no longer exists as such, I do not feel free to share the final reports for others to use and learn from. The reports may exist in a ‘public’ repository of some sort – perhaps the file drawer of a public servant who has taken over a programming domain and has yet to purge her files – but the effort and learning that went into and came from that evaluation process may not easily become part of a traditional meta-evaluation study. And such meta-evaluation work seems to be so rarely conducted, even though the sheer volume of evaluation work now being done in community-based organizations would seem to represent an untapped resource with enormous potential for learning and for informing deliberation. I suspect that such deliberative processes, if they led to influencing policy or even theoretical work, would provide a strong impetus for local organizations to participate, and to become part of efforts to change the world of programming for their clients and stakeholders.
Petrosino (2000) makes a fascinating observation and suggestion about the possibilities for meta-analysis in evaluation. He notes that the total amount of resources committed to evaluation work among all of the many active organizations and programs is enormous (p. 67). Even if we consider that many programs and agencies are not officially funded to do evaluation, the working time commitment of staff and volunteers alone is considerable. Petrosino notes that if the official budgets for such work were consolidated and a smaller sample of locations was studied in much more depth, the quality and potential consequence of such work could be potentially powerful and compelling. While I am not convinced of the appropriateness for devolved evaluation of the methods Petrosino recommends for his meta-evaluation approach, the idea of targeting specific forms of intervention for critical and comprehensive study is intriguing. Such a process might be done for comparable or fewer resources than are currently expended and still offer innumerable opportunities for social learning. But it would also depend on our ability to decouple evaluation from accountability – potentially a more challenging endeavour than promoting such meta-evaluation.

One other consideration that has linked numerous discussions throughout the dissertation is thinking about how we come to have the knowledge and practice the skills of the diverse roles of evaluation in a devolved context. I certainly appreciate that my own skills, for what they are worth, have been hard-won. I have made and continue to make many mistakes as I learn about representation, the nature of participation, what it means to facilitate deliberation, and to judge the receptiveness of contexts to any of these approaches. The traditional means by which people learn about doing evaluation through formal study make it challenging to discover such subtlety and nuance. Of necessity, I believe that it requires
immersion into complex contexts with all of their contingencies and ambiguity – the ‘rough ground’ that Schwandt (2003, 2006) talks about. Certainly it is easy to appreciate that devolved evaluation represents particularly rough terrain – such hills and valleys offer many unique learning opportunities. Having the opportunity to do such work under the supervision of a mentor, even as a volunteer working with a small grassroots or nonprofit organization, would expose learners to the challenges and accommodations necessary in real world evaluation, and I believe, would provide a lifetime’s worth of perspective on how rare those cases are where the perfect research design meets the willing and accommodating research setting, with enough resources, and adequate time to cover every interesting nook and cranny, leaving no (proverbial) stone unturned. I may yet see such a situation some day. Such an approach to teaching evaluation could have other benefits than simply pedagogical ones – addressing the capacity challenges within nonprofit organizations, drawing academia closer to community and linking both the ‘rough ground’ of practice with smooth theory, and perhaps demonstrating the similarities between devolved evaluation and more traditional conceptions of what and how evaluation work is done.

My reading of evaluation theory and practice is that we are already beginning a repositioning process in which mainstream evaluation is of necessity responding to the demands of devolved evaluation. By introducing more inclusive approaches to evaluation work, expanding the range of activities and roles undertaken by evaluators, and actively promoting more dialogical and democratically oriented evaluation emphases, evaluators are reshaping how the work is conceived and done. Yet the call for democratic approaches to evaluation, in particular, reflects a normative orientation that moves from the descriptive-
analytic mode to the prescriptive (Sork, 1996, p. 88), and the reality of many contexts is that they remain largely undemocratic. As Cervero and Wilson (2006) note:

… the major challenges to substantively democratic evaluation result from the asymmetrical political relationships in which the evaluation is designed and carried out. (p. 235)

This contradiction has been a part of my own efforts to understand and introduce more democratic approaches to evaluation in my work with nonprofit organizations. The context that demands that evaluation be done sets the parameters for what I can do in many, and perhaps most ways. It is a context that demands reflection on evaluation as a societal activity, with constraints and implications for broader governance, and the role of evaluation in that setting. Even so, my own emphasis – both within my working life and within the discussions in this dissertation – has also involved trying to understand and cope with those factors within the immediate control of evaluators such as myself; how we can change or adapt our practices to become more effective, and address those elements of the devolved context that are amenable to meaningful democratic action. This has led me to examine evaluation practice, and in particular devolved evaluation, as a site of leadership.

**Leadership in Evaluation Practice**

The evaluation literature has been virtually silent on the issue of leadership among evaluation practitioners. There is a substantial literature on evaluating leadership development programs, but this knowledge base about how to discover what works and does not work in ‘leadership’ has rarely been turned back on the discipline as a context of leadership. Similarly, while the management literature has had a strong emphasis in addressing leadership and the development of leadership, the emphasis has usually been on leadership within public and private sector organizations, and not on the practice of
consulting, or evaluation as a specialized sub-field within consulting practice. Leadership has also been studied in the nonprofit context, but usually not in the context of smaller grassroots organizations, and not in terms of evaluation, unless it is inferred that using or encouraging evaluation demonstrates leadership.

Much of the literature on leadership has emphasized leadership as a role that one takes on or a position to be filled. More recent literature has focused more broadly on the contexts of leadership, and following Terry (1997) I have come to see leadership as actions within specific contexts, and in particular, within contexts in which leadership is either required or appropriate. Terry (1998) makes the distinction between two aspects of context that are relevant to authentic leadership – certainty of outcome and level of agreement on direction. I find this distinction helpful in trying to understand how my own practice differs from that of other evaluators, particularly those working in a consulting role, or evaluators working with large corporate clients rather than small community-based projects. My work inevitably involves significant uncertainty of outcome, both for myself and for the organizations I work with and for. The broad diversity of interest groups in my work contexts, and the relative power differentials among players ensure that agreement on direction is likely to be low, and at the very least, subject to considerable negotiation. In contrast, in corporate settings, evaluation work typically has clearly identified and singular corporate clients, with the role and tasks of the evaluator carefully articulated in a contract as a technical expert rather than an advocate or mediator. When this type of evaluation work represented a major part of my practice, there was considerable negotiation to be undertaken with the work, but virtually all of it was with the client, rather than among competing interest groups or on behalf of client interests.
What really excites me is the leadership prospect for evaluators of using such tools to address some of the larger structural issues concerning how problems and solutions are framed. I have found that the devolution process – and perhaps it is a cultural message buried in how we frame community action – over-emphasizes the amount of control we have over context. It implicitly suggests that making fundamental societal changes is within the realm of possibility for those implementing short-term community-based projects. This denies the embeddedness of the problems to which the projects are addressed, and exaggerates the local power to effect change. Devolution thus can re-frame problems in terms of local solutions that are divorced from efforts at societal redress. I believe that evaluators can and must help to re-frame problems by clarifying the logical and actual linkages among programs, the capabilities of local efforts to address societal problems, and the role that local communities can and must play in that process. Evaluation can help frame what is locally contestable and what is not. Further, it can help address transformative issues – linking and bringing into the discourse those actors not in the immediate context, and either clarifying how and why some parts of the context are not negotiable, or bringing that context to the table as something that can be negotiated, given the proper participants and or power within the situation. Evaluators may be uniquely positioned to participate in addressing such concerns, or at the least, conducting meta-analysis of the contexts of social problems, solutions and policies. Building on the day-to-day leadership of evaluation practice through negotiation and facilitation, I think that the discipline is ripe for its own repositioning, as evaluators both act within context and act upon that context (Sork, 1996, p. 83).
Distinctive Contribution of the Study

Although I have had many opportunities to discuss the devolution of evaluation work with other evaluators and consultants at conferences, the process has not been extensively documented in the mainstream literature to date. Similarly, when evaluation in nonprofit organizations is examined, it is typically in larger organizations, it seldom reflects the context of local community agencies, and rarely addresses evaluation in grassroots organizations. As such, this study captures a broad and potentially far-reaching trend in a new way, and argues that it has profound implications for virtually all evaluation practice, most current evaluation practitioners, and a multitude of future evaluators in the grassroots.

The present study also contributes to recent efforts to examine and promote deliberative democratic evaluation and other collaborative evaluation approaches, including in particular participatory and developmental evaluation approaches. The emphasis on exploring the nexus of the expanded roles and functions of evaluation – including such topics as leadership and negotiation in evaluation practice – will be of interest to most evaluation practitioners. It consolidates a trend that is implicit in the work of many others, but for which the implications have not previously been explicitly articulated.

To those who are in the business of devolved program delivery, the topics of this study will hold little surprise, and yet they may provide some insight and possibly a reframe of the rationale for how they have been undertaking evaluation, and some of the potential benefits that evaluation can provide in supporting their ongoing efforts. Framing devolution in terms of evaluation and accountability complements the growing body of literature examining those trends, and hopefully extends it in useful ways.
Study Limitations

This study has focused on my reflections as an evaluation practitioner working as a consultant and volunteer in devolved program delivery settings. I have studied my own practice, and although I have discussed many of these issues with other practitioners and know that they have experienced similar pressures, I have not systematically gathered information about the practices of other evaluation consultants. Similarly, I know that a minority of grassroots and small nonprofit organizations actually hire consultants to assist them with their evaluation work. I have not examined evaluation conducted internally by such organizations, except in those situations in which I have been brought in to ‘fix’ an evaluation gone awry. Such evaluation work, performed by program staff or agency administrators will almost certainly differ substantially from the work that I undertake as an external evaluator, although the contexts with which they are coping (as discussed in Chapter Five, in particular) will be very similar.

Future Research

Like many studies, this one raises for me even more questions than it answers. One is a natural outgrowth of the way I have framed my questions and reflexively examined my practice – how common are my experiences? How have other consultants or evaluators been experiencing and reflecting on the process of devolution and how it is manifest in evaluation practice? A related question concerns all those nonprofit organizations that have not employed external evaluators as either consultants or volunteers – how are they coping with and experiencing the increased demands for evaluation that are part of the reality of program and service delivery in a world of devolution? For each of these questions I am interested in understanding both the experience and the meaning that it has for evaluators and those within
the organizations, and also gaining a better sense of the magnitude and scope of those changes. I can see the need for a mixed method research effort that could encompass detailed case studies of evaluation work with nonprofit organizations, and broad scale surveys of evaluators and nonprofit administrators about their experiences.

In my examination of the evaluation literature I have also come across several new and innovative approaches that would appear to hold promise for the kind of evaluation work I have been doing. Appreciative inquiry (Preskill & Catsambas, 2006) is one, as it seems to anticipate and actively work to reframe the expectations of evaluation in an upfront way that would focus it away from a hard accountability focus. Another is soft systems theory (Williams & Imam, 2007), which holds promise of helping to frame the diversity and complexity of nonprofit evaluation contexts in a way that might help identify important consistencies across cases and programs, and thus establish frame-factors relevant to meta-analysis and meta-negotiation.

In my work with nonprofit organizations I have been struck by the impacts of new information technologies on the expectations for reporting and accountability that they experience. This in itself would be a fruitful area for future research, as funders, umbrella organizations and individual nonprofit organizations work to establish automated information systems that will in turn frame their problems and solutions, and standardize their activities in potentially unanticipated ways.

Another issue implicit in my work to date is how gender affects all of these processes. Most of my contacts and the majority of the deliverers in local community agencies are women. Part of this can be understood culturally in terms of the way employment is structured in Canada, and the likelihood that women are filling positions that receive less
pay, less prestige and power in our society, and are in effect service organizations. I contrast this with my work with funders, who are at least as likely to be men as women, and this suggests questions that I believe are worth pursuing in more detail: ‘How does gender affect the framing of evaluations, evaluation questions, and the approaches used in devolved evaluation?’ ‘What are the implications of the gendered nature of work in nonprofit and grassroots organizations for how and how much evaluation is demanded and conducted?’

Another observation flowing from my work and this study is the way that resistance to the imposition of power, accountability demands and control over the work of nonprofit organizations has developed and takes place. Indeed, many of my contacts have expressly noted and described the evaluation function as part of the cost of living in a surveillance society, and accept it grudgingly, but with active and usually subtle forms of resistance. Most of my contacts and virtually all devolved delivery settings involve active negotiation of these power relationships. The form, extent and effectiveness of this resistance are relevant to better understanding the potential need for and activities of evaluators as negotiators and leaders. This could be an enormous and important area of future research on nonprofit organizations.

Finally, my own interests in the policy context of program and service delivery have been enhanced by seeing the dance over advocacy and contributions to government policy that are both implicit and explicit in devolved delivery. This process seems to be a growing area of contestation, and the role of evaluation in contributing to policy discussions and helping to lever the participation of the nonprofit sector in such discussions lies at the heart of where I believe efforts to introduce horizontal governance are headed. I have alluded to policy as an area of great concern to local organizations and relevance to evaluators, and the
nature of inclusiveness in policy discussions represents another extremely important part of the devolution puzzle. And as Walden (2006) notes, the growing culture of surveillance that surrounds nonprofit governance also demands that nonprofit organizations proactively strategize about how to challenge limitations on advocacy and risk taking with respect to program innovation and policy contribution.

Reflections on Researching my Practice

I began my evaluation career having taken no formal instruction in evaluation, other than a brief mention in traditional undergraduate research methodology courses. I had the fortunate opportunity to learn from a mentor – someone who had both studied the theory and practice of evaluation, and had conducted numerous studies in diverse contexts. My first lesson in evaluation consisted of an hour-long briefing as preparation immediately prior to departing for a Canada-wide tour in which I conducted focus groups as part of a national evaluation study. I have strong memories of my trial by fire as a young and inexperienced researcher scrambling to conduct credible focus group interviews with wary program deliverers and participants, but my main memory of that whirlwind briefing is the heartfelt advice offered: “Remember that context is everything, and that you will need to be sensitive, flexible, and responsive to the political and social setting you find in each community.”

Because of this, as I began to read about evaluation and learn about it through more formal study, I was particularly appreciative of Cronbach’s work (1982), in which he describes designing an evaluation as both a science and an art, and says that every program evaluation is a unique undertaking (pp. 1, 321). His final word on the rationale for his book is that such a book on evaluation is fundamentally contradictory; that those who could potentially benefit most from his book are those most familiar with what it has to say.
No one learns to plan and conduct evaluations from a book. The way to learn is to try and try again, part of the time under circumstances where a more experienced practitioner can comment on the plans and suggestions that emerge from the early data. (p. 335)

Given my own initiation to evaluation practice thanks to a conscientious mentor, I am aware of the debt I owe to the many patient individuals – many of them subjects of the studies I have undertaken and participated in – who have over the years provided me with the benefit of their own experience and knowledge. In my recent consulting practice with nonprofit organizations I have found some ways to pay back to communities and organizations through my work, by educating people about doing evaluation. As I have recently made another transition to a teaching role at a university, I hope to be able to continue this work by actively mentoring student researchers, while continuing my work with local community agencies as both a consultant and a volunteer. Doing so addresses some of my key concerns with the nature of evaluation work in a devolved setting – finding ways to accommodate the significant capacity challenge of such organizations, expanding the benefits to agencies that have not been able to afford a ‘consultant’ to help with the work, and further developing my own ability to provide the linking, bridging, negotiating and framing work that is so needed in the sector. For most grassroots and small nonprofit organizations, evaluation work is still done off the sides of their desks – and they need to find a way to accomplish the work, and perhaps a desk for it to rest on. For some of them, at least, it may as well be mine. I can try to keep it from falling off the desk altogether.
REFERENCE LIST


http://www.europeanevaluation.org/docs/ArnkilSeville200220.pdf


# APPENDICES

## Appendix A: Meta-Analysis Tables

### TABLE A1: DEVOLUTION CASE STUDIES AND SURVEY ARTICLES

<table>
<thead>
<tr>
<th>Article lead author for case studies on devolution</th>
<th>Devolution Meta-Themes</th>
<th>Evaluation Meta-Themes</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Accountability</td>
<td>Capacity</td>
</tr>
<tr>
<td>CASE STUDIES (+ Surveys)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Akingbola (2004)</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Alexander, et al. (1999) (+S)</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Alexander (1999) (+S)</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Alexander (2000) (+S)</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Brilliant &amp; Young (2004)</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Burnley, et al. (2005) (+S)</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Campbell (2002)</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>DeVerteuil, et al. (2002)</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Ebrahim (2002)</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Levasseur &amp; Phillips (2005)</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Marwell (2004)</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Mullett, et al. (2002)</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Mulroy (2003)</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Mulroy &amp; Tamburo (2004)</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Tassie, et al. (1998)</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>SURVEYS</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Considine (2002)</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>McMurtry, et al. (1991)</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Rawsthorne (2005)</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Sommerfeld &amp; Reisch (2003)</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Swindell (2000)</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

Legend for Tables A1 & A2:

- ✓ Key emphasis
- ✓ Relevant emphasis
<table>
<thead>
<tr>
<th>Article lead author for case studies on evaluation</th>
<th>Devolution Meta-Themes</th>
<th>Evaluation Meta-Themes</th>
</tr>
</thead>
<tbody>
<tr>
<td>CASE STUDIES (+ Surveys)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Abma (2000b)</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Abma (2007)</td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Baulderstone (2005)</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Boser (2002)</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Campbell, et al. (2003)</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Clayson, et al. (2002)</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Costantino &amp; Greene (2003)</td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>English &amp; Kaleveld (2003)</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Folkman &amp; Rai (1997)</td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Fredericks, et al. (2002)</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Freedman (1998)</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Graham &amp; Bois (1997)</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Greene (2002b)</td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Huebner (2000)</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>MacNeil (2002)</td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>MacNeil (2000)</td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Mathison (2001)</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Mathison (1994)</td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Minnett (1999)</td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Morabito (2002)</td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Ryan &amp; DeStefano (2000)</td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Themessl-Huber &amp; Grutsch (2003)</td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Wallerstein (1999)</td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Westley et al. (2006)</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>SURVEYS</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Carman (2007)</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Carman &amp; Millesen (2005)</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Cousins, et al. (1996)</td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Fine, et al. (2000)</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>
### Appendix B: Canadian Nonprofit & Voluntary Organizations

**TABLE B1**  
**NONPROFIT AND VOLUNTARY ORGANIZATIONS, BY PRIMARY ACTIVITY AREA, CANADA, 2003**

<table>
<thead>
<tr>
<th>Primary activity areas</th>
<th>Includes...</th>
<th>Number of organizations</th>
<th>% of all organizations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arts and culture</td>
<td>organizations and activities, whether they are in general or specialized fields of arts and culture</td>
<td>13,770</td>
<td>8.5</td>
</tr>
<tr>
<td>Sports and recreation</td>
<td>organizations and activities, whether they are in general or specialized fields of sports and recreation</td>
<td>33,649</td>
<td>20.9</td>
</tr>
<tr>
<td>Education and research</td>
<td>organizations and activities administering, providing, promoting, conducting, supporting and servicing education and research, excluding public school boards, universities and colleges</td>
<td>8,284</td>
<td>5.1</td>
</tr>
<tr>
<td>Universities and colleges</td>
<td>postsecondary institutions</td>
<td>502</td>
<td>0.3</td>
</tr>
<tr>
<td>Health</td>
<td>organizations that engage in health-related activities, provide and administer both general and specialized health care services, as well as health support services, excluding hospitals</td>
<td>5,324</td>
<td>3.3</td>
</tr>
<tr>
<td>Hospitals</td>
<td>hospital facilities that provide inpatient or outpatient medical care</td>
<td>779</td>
<td>0.5</td>
</tr>
<tr>
<td>Social services</td>
<td>organizations and institutions providing social services to a community or target population</td>
<td>19,099</td>
<td>11.8</td>
</tr>
<tr>
<td>Environment</td>
<td>organizations promoting and providing services in environmental conservation, pollution control and prevention, environmental education and health, and animal protection</td>
<td>4,424</td>
<td>2.7</td>
</tr>
<tr>
<td>Development and housing</td>
<td>organizations promoting programs and providing services to help improve communities and promote the economic and social well-being of society</td>
<td>12,255</td>
<td>7.6</td>
</tr>
<tr>
<td>Law, advocacy and politics</td>
<td>organizations and groups that work to protect and promote civil and other rights, advocate the social and political interests of general or special constituencies, offer legal services and promote public safety</td>
<td>3,628</td>
<td>2.3</td>
</tr>
<tr>
<td>Grant-making, fundraising</td>
<td>philanthropic organizations and organizations promoting charity and charitable activities including grant-making foundations, voluntarism promotion and support, and fundraising organizations</td>
<td>15,935</td>
<td>9.9</td>
</tr>
<tr>
<td>and voluntarism promotion</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>International</td>
<td>organizations promoting cultural understanding between peoples of various countries and historical backgrounds, as well as those providing relief during emergencies and those promoting development and welfare abroad</td>
<td>1,022</td>
<td>0.6</td>
</tr>
<tr>
<td>Religion</td>
<td>organizations promoting religious beliefs and administering religious services and rituals (such as churches, mosques, synagogues, temples, shrines, seminaries, monasteries and similar religious institutions); and related organizations and auxiliaries of such organizations</td>
<td>30,679</td>
<td>19.0</td>
</tr>
<tr>
<td>Business and professional</td>
<td>organizations promoting, regulating and safeguarding business, professional and labour interests</td>
<td>8,483</td>
<td>5.3</td>
</tr>
<tr>
<td>associations and unions</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Organizations not</td>
<td>a small number of organizations that were unable to locate their primary activity within any of the specified activity categories</td>
<td>3,393</td>
<td>2.1</td>
</tr>
<tr>
<td>elsewhere classified</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>All organizations</strong></td>
<td></td>
<td><strong>161,227</strong></td>
<td><strong>100.0</strong></td>
</tr>
</tbody>
</table>

**Note:** Figures do not add to total and percentages do not add to 100 due to rounding.

**Data Source:** Statistics Canada, Survey of Nonprofit and Voluntary Organizations.

**Table source:** Statistics Canada, 2004. *Cornerstones of Community*, Table 1.1, p. 13.

---

*Statistics Canada information is used with the permission of Statistics Canada. Users are forbidden to copy the data and redisseminate them, in an original or modified form, for commercial purposes, without permission from Statistics Canada. Information on the availability of the wide range of data from Statistics Canada can be obtained from Statistics Canada’s Regional Offices, its World Wide Web site at www.statcan.ca, and its toll-free access number 1-800-263-1136.*

---

341