THE INFLUENCE OF EVALUATION ON HUMAN SERVICE PRACTITIONER LEARNING: AN ANALYSIS OF TWO CASE STUDIES OF PRACTITIONERS' EXPERIENCES OF EVALUATION

James Leslie Herbert
Bachelor of Arts (Psychology)
Master of Criminal Justice

This thesis is presented in fulfilment of the requirements for the degree of Doctor of Philosophy

School of Social Sciences and Psychology
University of Western Sydney

July, 2013
DEDICATION

This thesis is dedicated to the memory of my father Bruce Stephen Herbert, a quiet and thoughtful man with a lifelong love of learning. The older I get, the more I realise I am like him in so many ways, and that I can continue to know him through myself. The best parts of me are him.
ACKNOWLEDGEMENTS

This thesis would not have been possible without the guidance of supervisors Mick Houlbrook, Natalie Bolzan, and Wayne Fallon. I cannot thank them enough for their support and for their unwavering confidence in me. I acknowledge Margaret Mitchell my early supervisor who got me to the confirmation of candidature stage. I also acknowledge Romy Lawson, an important mentor in my continuing academic development. I am very grateful to both the organisations I conducted my research within, but I would like to particularly thank Andrew Anderson, Saul Flaxman and Sally Cowling for their assistance. I am also similarly grateful to the participants who were very generous with their time and experiences. Though distant in geography, my family are a never ending source of support for me; I thank my mother Christine Herbert, brother Matthew Herbert, and sister Stephanie Herbert. I thank friends who have helped make Sydney a bit more like home: Braith, Nicky, Natalie, Megan, Gabe, Lauren, Mark, Rebekah, Jason, Emese, Keith, Rebecca, Wayne, Meredith, Matt, James, and Sarah; and friends back home who I miss: Nez, Raffe, Amy, Earl, Vic, Courtney, Jen. Finally I thank the other post-graduates who I have shared time, space and ideas with: Guy, Deborah, Sera, Geir, Annette, Allesandro, Gordon, Rosie, Ashley, and Miriam.
STATEMENT OF AUTHENTICATION

The work presented in this thesis is, to the best of my knowledge and belief, original except as acknowledged in the test. I hereby declare that I have not submitted this material, either in full or in part, for a degree at this or any other institution.

.................................................................
(Signature)

.................................................................
(Date)
# TABLE OF CONTENTS

Dedication.................................................................................................................. ii
Acknowledgements .................................................................................................. iii
Statement of Authentication ...................................................................................... iv
Table of Contents..................................................................................................... v
List of Tables ........................................................................................................... ix
List of Figures .......................................................................................................... x
Abbreviations ........................................................................................................... xi
Abstract .................................................................................................................. 1

1. Chapter 1: Introduction and the Context of Human Service Provision...... 3
   1.1 Introduction ....................................................................................................... 3
   1.2 Thesis Background ......................................................................................... 4
   1.3 Research Problem .......................................................................................... 7
   1.4 Aims and Scope ............................................................................................. 7
   1.5 Implications and Significance ........................................................................ 8
   1.6 The Context of Human Service Provision ...................................................... 9
      1.6.1 NGOs as the Primary Provider of Human Services ............................... 10
      1.6.2 Neo-Institutionalisation as a Colonising Force ................................... 15
      1.6.3 The Perceived Marginalisation of Human Service Practice ............... 16
   1.7 The Challenge for Evaluation ........................................................................ 20
   1.8 Overview of Thesis Structure ...................................................................... 21
   1.9 Chapter Summary .......................................................................................... 23

2. Chapter 2: Practicing Evaluation for Learning and Improvement ............ 25
   2.1 Introduction ....................................................................................................... 25
   2.2 Defining Evaluation ....................................................................................... 26
      2.2.1 Values in Evaluation .............................................................................. 28
   2.3 What is the Intended Use of Evaluation? .......................................................... 30
   2.4 How Evaluation Use is Conceptualised in the Literature ............................. 32
      2.4.1 Use/Utilisation: 1970-1986 .................................................................... 32
      2.4.2 Evaluation Use/Impact: 1986-2000 ......................................................... 37
      2.4.3 Evaluation Influence: 2000-Present ......................................................... 40
   2.5 The Evaluation Influence Frameworks ............................................................. 45
   2.6 Evaluation Influence Research ....................................................................... 47
      2.6.1 Case Studies ............................................................................................. 48
      2.6.2 Surveys .................................................................................................... 50
6.5 The Utility of Evaluation Influence in Describing Practitioner Learning ...........................................................240
6.6 Limitations of the Research ....................................................................................................................................245
6.7 Chapter Summary ......................................................................................................................................................246

7. Chapter 7: Conclusion ...........................................................................................................................................248
7.1 Introduction ............................................................................................................................................................248
7.2 Key Findings ............................................................................................................................................................253
  7.2.1 The Relationship between Evaluation and Practice ..........254
  7.2.2 The Validity of Evidence to Inform Practice ..........254
  7.2.3 Equitable and Participatory Engagement ..............255
  7.2.4 Acknowledging the Critical Politics of Evaluation ..........255
  7.2.5 Resolving Potential Conflict between Inter-Organisational Priorities .................................................................256
7.3 Recommendations ....................................................................................................................................................256
  7.3.1 Commitment to Practice-Oriented Evaluation from Funding Bodies .................................................................256
  7.3.2 Demonstrate Organisational Commitment to Practitioner Learning from Evaluation ....................................257
  7.3.3 Conduct Evaluation in a Way that Enhances Practice Influence ......257
  7.3.4 Continuing Research into How to Overcome the Challenges of Integrating Research and Evaluation in Practice ................................................................................................................258
7.4 Future Research .........................................................................................................................................................259
7.5 Conclusion .................................................................................................................................................................259
LIST OF TABLES

Table 3.1 Ranked Factors Associated with the Use of Evaluations .................. 71
Table 3.2 Additional Factors Associated with the Use of Evaluations .................. 72
Table 3.3 Factors that Increase Evaluation Use ........................................... 75
Table 3.4 Factors Associated with Evaluation Use in NGOs .......................... 78
Table 4.1 Description of Participant Groups .............................................. 109
Table 4.2 The Benevolent Society Participant Details .................................. 112
Table 4.3 The Benevolent Society Organisational Documents included in Analysis ................................................................. 113
Table 4.4 The Benevolent Society Timeline of Significant Events ................. 113
Table 4.5 Burnside Participant Details ....................................................... 114
Table 4.6 Burnside Organisational Documents included in Analysis ............ 115
Table 4.7 Burnside Timeline of Significant Events ..................................... 115
Table 4.8 Definitions of the Individual Mechanisms of Evaluation Influence .... 122
Table 4.9 Definitions of the Interpersonal Mechanism of Evaluation Influence . 127
Table 4.10 Collective Mechanisms of Evaluation Influence .................................. 131
Table 5.1 Local Government Areas where The Benevolent Society Delivers Brighter Futures ................................................................. 143
Table 5.2 Mechanisms Reported in the Brighter Futures Evaluation .............. 152
Table 5.3 Mechanisms Reported in the St Marys Evaluation .......................... 191
LIST OF FIGURES

Figure 2.1 Owen’s (2007) Two Definitions of Evaluation ........................................ 27
Figure 2.2 A Comparison of Evaluation Influence Frameworks .......................... 46
Figure 3.1 Evaluation Implementation Factors to Explore in Practitioner Learning
from Evaluation ........................................................................................................... 81
Figure 3.2 Stakeholder Involvement Factors to Explore in Practitioner Learning from
Evaluation .................................................................................................................. 82
Figure 3.3 Practice or Organisational Setting Factors to Explore in Practitioner
Learning from Evaluation ............................................................................................ 83
Figure 4.1 Timeline of the Evaluation in Relation to the Case Study Period ........... 107
Figure 4.2 Data used in Each Phase of the Analysis ................................................. 119
Figure 6.1 Conditions and Barriers to Practice Learning from Evaluation ............. 216
### ABBREVIATIONS

<table>
<thead>
<tr>
<th>Acronym</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABC</td>
<td>Australian Broadcast Company</td>
</tr>
<tr>
<td>ACOSS</td>
<td>Australian Council of Social Services</td>
</tr>
<tr>
<td>ACTCOSS</td>
<td>Australian Capital Territory Council of Social Services</td>
</tr>
<tr>
<td>AEA</td>
<td>American Evaluation Association</td>
</tr>
<tr>
<td>AES</td>
<td>Australasian Evaluation Society</td>
</tr>
<tr>
<td>ARC</td>
<td>Australian Research Council</td>
</tr>
<tr>
<td>BF</td>
<td>Brighter Futures</td>
</tr>
<tr>
<td>CaLD</td>
<td>Culturally and Linguistically Diverse</td>
</tr>
<tr>
<td>COAG</td>
<td>Council of Australian Governments</td>
</tr>
<tr>
<td>DOCS</td>
<td>Department of Community Services (NSW)</td>
</tr>
<tr>
<td>FaHCSIA</td>
<td>Families, Housing, Community Services and Indigenous Affairs</td>
</tr>
<tr>
<td>INGO</td>
<td>International Non-Government Organisation</td>
</tr>
<tr>
<td>MADD</td>
<td>Mothers Against Drink Driving</td>
</tr>
<tr>
<td>NEWPIN</td>
<td>New Parent and Infant Network</td>
</tr>
<tr>
<td>NGO</td>
<td>Non-Government Organisation</td>
</tr>
<tr>
<td>OECD</td>
<td>Organisation for Economic Cooperation and Development</td>
</tr>
<tr>
<td>PIP</td>
<td>Performance Improvement Plan</td>
</tr>
<tr>
<td>SPRC</td>
<td>Social Policy Research Centre</td>
</tr>
</tbody>
</table>
ABSTRACT

Privatisation and the contracting out of human services to Non-Government Organisations (NGOs) is a feature of most countries with a strong ideological connection to neo-liberalism. The outsourcing of these services is accompanied with systems of performance management, along with program evaluation. Contemporary program evaluation emphasises the importance of collaborative and participatory processes to engage stakeholders, which enables program improvement, yet typically these evaluations seem to address the narrow accountability needs of service funders (Ebrahim, 2005a). Given that NGOs by and large have significant commitments to evidence based practice (Carman & Fredricks, 2008), and the improvement of program practice is in the interests of all parties, what are the conditions and barriers for practitioner learning from evaluation? This research involves two case studies of evaluations that aimed to foster practice learning, these will be explored in order to develop an understanding of the conditions and barriers to practitioner learning from evaluation.

This research is novel in that it focuses specifically on practice level learning, an individually oriented phenomena that has received limited focus in the evaluation literature. The thesis draws on evaluation influence\(^1\) (Mark & Henry, 2004) as a framework to explore the broad and diffuse impact that evaluations can have, an approach well suited to exploring learning. Evaluation influence presents a series of mechanisms at the individual, interpersonal, and collective levels that reflect the full impact of evaluation and a cohesive way to organise theoretical and empirical knowledge of evaluations. Drawing on an understanding of the case, and the

\(^1\) The terms ‘evaluation influence’ and ‘evaluation use’ have been italicised throughout the document to highlight the fact that these terms are conceptualisations of the impact an evaluation can have.
identification of influence mechanisms, the research sought to develop an understanding of the conditions and barriers to practitioner learning from evaluation.

Analysis of these case studies suggests that there seems to be five key issues related to practitioner learning from evaluation. There was a lack of clarity in the relationship between evaluation and practice, despite there being organisational commitments to evidence based practice. There was a need to produce evaluation that had the validity to inform practice, evaluation with practice credibility as well as the kinds of validity and reliability understood by evaluators. Influence comes from the equitable and participatory engagement of practitioners in the evaluation, valuing their input and treating practitioners as stakeholders that possess important program and implementation knowledge. Acknowledging the critical politics of evidence was important in representing the practice context, and assuring stakeholders that the evaluation was not merely symbolic. Finally, managing the clash of inter-organisational priorities was imperative to getting practice learning on the agenda of an evaluation.

Practice oriented evaluation (Schwandt, 2005) has the potential for significant benefits that translate to improvements in service delivery to isolated and vulnerable communities. This thesis adds to the concept of practice oriented evaluation by suggesting a number of important elements that it should include: embedding evaluation into practice; creating evidence that is valid to inform practice; making effective use of practitioners as evaluation stakeholders; engaging with the critical politics of organisations and the sector; and advocating for practice oriented evaluation at the planning stages of an evaluation.
1. CHAPTER 1: INTRODUCTION AND THE CONTEXT OF HUMAN SERVICE PROVISION

1.1 Introduction

The first chapter of this thesis sets out with two main objectives. First, to direct the reader to clearly understanding and connecting the sections and chapters that will follow. Secondly to provide a context for the thesis. The chapter will begin by briefly providing some background to defining the research problem, setting out the aims and scope of the research, the implications and significance, and finally an overview of the structure of the thesis that follows. As a precursor to chapter two, the second part of this chapter seeks to provide the reader with a sense of context to ground the research in the broader issues of the sector.

International trends have shaped the changing relationship between Non-Government Organisations (NGOs) and the funders of social services (Bourdieu, 1998). There are broad concerns from the human service practitioner community about how these developments impinge on the ability of workers within NGOs to use professional discretion to practice in a way that is responsive to clients (Baines, 2011; Johnson-Abdelmalik, 2011; Meagher & Healy, 2003; Meagher & Parton, 2004). Providing some detail about these trends starts the process of connecting these broad issues to the day-to-day challenges and experiences of human service practitioners. Considering these trends will lead into seeking to understand human service practitioners’ experiences and perceptions of evaluation that take place within this socio-political context.
1.2 Thesis Background

Privatisation and the contracting out of services traditionally provided by governments has become a feature of most countries in the Organisation for Economic Co-operation and Development (OECD, 2009). Neo-liberal economic discourses are an international phenomena informing these developments (Bourdieu, 1998). Australian governments have been particularly active in contracting out non-core government activities (e.g. telecommunications, utilities, transport and logistics; Schneider, 2003). In the welfare and human services sector NGOs have quickly become the primary service providers in Australia (Productivity Commission, 2010), with other developed countries following suit (Wright, Marston, & McDonald, 2011). As NGOs increasingly interact with and are accountable to government organisations, neo-institutional theory suggests that NGOs are likely to operate more like government bureaucracies and privilege the same values as these organisations (Di Maggio & Powell, 1983; Meyer & Rowan, 1991). The impact of the shift to the purchasing of social services (O'Shea, 2007), the emergence of private sector funding models (The Centre for Social Impact, 2011), and the legitimacy of appearing businesslike under New Public Management (Goerke, 2003) are powerful influences on the structure and values of NGOs. These shifts are dramatic recent developments in a sector that has previously remained staunchly independent of government and private sector influence, indeed the description of NGOs as the third sector refers to their conscious separation from politics and markets (Wolfe, 1991). Amongst staff who implement the goals of the organisation through their interactions with clients, these broad changes are perceived as the unwanted incursion of managerialism and bureaucracy into the sector (Baines, 2011; Johnson-Abdalmalik, 2011; Meagher &
Healy, 2003; Meagher & Parton, 2004; Singh & Cowden, 2009), characterised by a preoccupation with restrictive types of accountability (Ebrahim, 2005a).

Practitioner groups often see accountability measures such as evaluation and performance measurement as narrowly serving the interests of funders through pseudo-scientific and technocratic measures. Evaluation is criticised as being reductive in the way it represents human service work (Mintzberg, 1996; Sanderson, 1998; Schwandt, 2005; Shaw & Falkner, 2006), often threatening to the continuation of the program (Donaldson, 2002; Kegeles, Rebchook, & Tebbetts, 2005), and at best irrelevant to the real work of helping service users (Meagher & Healy, 2003). Yet, despite this sense of being part of the managerialist and technocratic movement, evaluation has been on a trajectory of change. Early program evaluation was much more like a kind of applied social science research, attempting to produce objective and scientific information to inform decision-makers (Owen, 2007). Responding to early research that suggested evaluation had limited influence on decision-making (Alkin, Kosecoff, Fitzgibbon, & Seligman, 1974; Caplan, Morrison, & Stambaugh, 1975; Patton et al., 1977), evaluators conceptualised a broader role for evaluation. Stemming in part from Patton’s (1978) seminal text *utilisation-focused evaluation*, evaluation practice began to reflect program context (Shulha & Cousins, 1997). Engaging with context emphasises understanding process alongside outcomes (Patton, 1998), collaborative approaches to evaluation (OECD, 2009), and evaluation capacity building within organisations (Schneider, 2003), each of which have become important parts of contemporary evaluation practice. Improvement and learning are viewed as important outcomes of these approaches to evaluation, particularly at the practice level where learning can mean improvements to the quality of services and to the lives of service users (Schwandt, 2005). This thesis is
concerned with the challenge of getting substantive improvements at the practice level through practitioner learning from evaluation.

Defining learning derived from evaluation presents a challenge due to the trans-disciplinary nature of contemporary evaluation (Scriven, 2003). While human services are delivered in the framework of a program, the workers that implement program aims often have considerable autonomy and discretion (Evans & Harris, 2004). Organisational learning systems can retain and disseminate information (Daft & Weick, 1984), yet due to the level of discretion these worker have, organisational learning is limited in describing practice change. This thesis recognises that changes in practice occur from “the liberation of knowledge from self-reflection and questioning” (Antonacopoulou, 2001, p. 328), which are prompted by change processes at the individual, group, or organisational level. This emphasises that, for human service practitioners, change is primarily an individually oriented process, however taking place in this broader context of information and value exchange.

This thesis focused on the process by which evaluation influence (Henry & Mark, 2003; Kirkhart, 2000; Mark & Henry, 2004) results in practitioner learning from evaluation in NGOs, in order to develop an understanding of the conditions and barriers to this happening. Evaluation influence represents a systematic framework of mechanisms that describe the way an evaluation can affect change (Mark & Henry, 2004). While the term evaluation use is still common in the literature, evaluation influence has developed around the failure of ‘use’ to reflect the full effects an evaluation can have (Herbert, 2011). This broader conceptualisation of the effects of evaluation makes evaluation influence ideal for the exploration of practitioner learning from evaluation.
1.3 Research Problem

This research is focused on practitioner learning from evaluation. There are inherent challenges in advancing the aim to foster this type of influence. While surveys of evaluators show that learning is an important outcome of an evaluation (Carman & Fredricks, 2008), evaluations still tend to be oriented towards providing upwards accountability (Carman, 2007; Ebrahim, 2005a; Hoole & Patterson, 2008). The complex socio-political environment of human service organisations can confound evaluation practice even when evaluators are mindful of practitioner learning. Efforts to provide human service practitioners with useful feedback about their work is situated in a wider milieu of contestation about the role of NGOs as service providers and their relationship with government funders (Maddison, Denniss, & Hamilton, 2004; O'Shea, 2007). Added to the traditional unfamiliarity of evaluation procedures in the NGO sector (Taut & Alkin, 2003), there are broad institutional challenges in fostering learning from evaluation.

The questions dealt with by this thesis are:

- What are the conditions and barriers to human service practitioner learning from evaluation in human service programs delivered by NGOs; and
- How could evaluation be undertaken in order to maximise practitioner learning?

1.4 Aims and Scope

This thesis sets out with a number of interrelated aims. The initial aim is to obtain a comprehensive understanding of the staff reaction to two evaluations that took place within human service programs run by NGOs. Drawing on interviews and organisational evaluation documents across different levels of the organisations over the course of a year, this understanding will arise from the collation of these pieces of information into a narrative of the evaluation. The second aim concerns the use of
Mark & Henry’s (2004) *evaluation influence* framework in order to identify where the evaluation has affected change at the individual, interpersonal, and collective levels. This framework provides a set of change mechanisms at these different levels that allow for the exploration of the broad and diffuse impacts evaluation can have. The third aim of the research is to draw on instances of *evaluation influence* and the participants’ experiences of how influence occurred in order to infer the conditions and barriers to human service practitioners’ learning from evaluation. These will be developed into a set of insights that have clear implications for evaluation practice. This research will add to the existing knowledge of *evaluation influence* and evaluation impact more broadly, and is novel in focusing on the use of evaluation by human service practitioners rather than decision-makers.

The scope of this thesis includes a number of components: (a) practitioner learning, which is thoughtful processing and reflection on evaluation information that leads to changes in practice (Antonacopoulou, 2001; Fleming, 2011); (b) evaluation, meaning program evaluation, a body of literature that has developed around the challenges of assessing the value or worth of publicly funded programs (Owen, 2007); (c) human service programs, where the service is constituted by the human interaction between a worker and client; and (d) NGOs, as organisations that are traditionally distinct from politics and the market (Productivity Commission, 2010; Wolfe, 1991).

### 1.5 Implications and Significance

By exploring the conditions and barriers for evaluation influencing practitioners this thesis lays the groundwork for an approach to evaluation that pursues social betterment (Mark, 2001). This approach strongly encourages the engagement of staff who are involved in the difficult daily decisions associated with implementing social
Evaluation for Practitioner Learning  

Chapter 1: Introduction and Context

programs. Evaluation is likely to become an increasingly important part of human services as governments seek to divest themselves of service provision, and take on the role of purchasers of social services (O'Shea, 2007; O'Shea, Leonard, & Darcy, 2007). While evaluations can produce findings that can affect positive change, improving service delivery directly is arguably the most effective means of achieving change in the experiences and outcomes of service users. This thesis argues that contracting, designing, and conducting evaluation that has practice relevance represents an important step in advancing the broader legitimacy of evaluation.

1.6 The Context of Human Service Provision

This section aims to position the discussion about practitioner learning from evaluation in the broader context of human service provision in Australia. For the purpose of this thesis the term human services refers to a broad subset of social, community, and welfare oriented work, whose practitioners include social workers, psychologists, case workers, and facilitators (Australian Council of Social Services, 2011). While a number of terms are available to describe this work and these workers, the term human services has been preferred as it describes the core facet of the work, interactions with clients which are emotional, personal, and above all human (Kirkland, 2001).

This kind of work is inevitably personal and emotional (Kirkland, 2001), and research suggests that the chance to live and express personal values in work is a significant determinant of the high levels of job satisfaction in the industry (Wright, et al., 2011). Given the centrality of values and professional autonomy to human service work, it is particularly important to get a sense of broader trends in order to better understand and interpret their impact at the organisational and practice level. Also of importance is the shift in Australia to NGOs as primary service providers, a
change that continues to alter the meaning of human services and their practice. In
discussing the context of human service provision, the aim is to draw out some of the
forces for change and effects that these changes have had on human service
organisations and their work. Evaluation features as part of these changes, both as a
force itself and as part of the trend for increased accountability in the sector
(Ebrahim, 2005b).

1.6.1 NGOs as the primary provider of human services
Since the 1980s, governments worldwide have been in the process of divesting
themselves of responsibility for direct service delivery, prompted primarily through
an ideological commitment to neo-liberalism (Bourdieu & Wacquant, 2001). Neo-
liberal economic thinking is a global trend (Bourdieu, 1998) oriented towards
reducing the involvement of the state in civil society. The argument follows that
governments acting as monopolistic producers and distributors are inefficient
because of higher salaries, political patronage, unions, red tape, and a lack of
incentive for innovation (Wolfe, 1991). In Australia, the National Competition Policy
launched by the Council of Australian Governments (COAG, 1995) signalled the
intention for the market to take over all non-core or non-legislated service delivery.
While academic debate about the advantages and disadvantages of privatisation have
raged (Aulich & O'Flynn, 2007; King, 1998), the process has rolled on with total
government funding to NGOs increasing from $10.1 billion in 1999-2000 to $25.5
suggests that neo-liberalism has been elevated to a hegemonic status in public
discourse, an implicit and irrefutable understanding with no credible alternative.
Human services in particular have been altered by neo-liberal discourses, not just in
the way they are funded and delivered, but also in terms of what they are understood
to be for.
In a relatively short period of time human services have undergone dramatic change, primarily the disengagement of governments from direct service delivery, prompted by the re-conceptualisation of such services as no longer being part of social citizenship rights (Ziguras, Dufty, & Considine, 2003). Welfare and human service provision have long been acknowledged in public policy as a way to manage the social costs of private enterprise (Jansson, 2005), but were in the 1970s beginning to be presented as a set of social democratic rights (McDonald & Reisch, 2008). This sense of rights has been eroded by the effects of neo-liberalism, commitment to which has been particularly strong in Australia (Jamrozik, 2001). Replacing this sense of collectivism and of human services being a right, is the view of service users as customers who are assumed to have an unfettered choice in the types of services they can access (Garland & Darcy, 2009). The disengagement of government from human service provision reflects the preference for a market of services (Productivity Commission, 2010), provided primarily by NGOs.

NGOs are also often referred to in the literature as non-profit organisations, community organisations, or the third sector (Productivity Commission, 2010), with slight variations on their definition. While the term NGO is often associated with international aid, for the purposes of this thesis it reflects a broad definition describing a diverse sector (Najam, 1996). This research is concerned specifically with human service NGOs, so organisations involved in the delivery of services to the community, often to vulnerable client groups, in traditionally non-market settings (Productivity Commission, 2010). NGOs are characterised as being driven by mission rather than profit, with their broader support in the community associated with their efforts in pursing this community purpose (Productivity Commission, 2010). While in other sectors service roles have been taken on by for-profit
providers, welfare and human services have increasingly been contracted out to NGOs (Kirkland, 2001). This has taken place for a number of reasons, but primarily due to lower service costs associated with: not having to make a profit (Lister, 2003), the receipt of donations and in-kind volunteering (Productivity Commission, 2010), lower wages for staff compared to government workers (Australian Capital Territory Council of Social Services, 2008; Western Australian Council of Social Service, 2009), and significantly lower overheads related to administration and bureaucracy (Bedsworth, Gregory, & Howard, 2008). Van Slyke and Roch (2004) suggest that NGOs are attractive recipients of government contracts as they have less of an incentive to exploit the information dissymmetry between themselves and funders in order to pursue profits.

Beyond lower service costs, NGOs enjoy a number of advantages as providers of social services. There is evidence to suggest that despite the lower pay and conditions, NGO staff are a highly motivated and dedicated workforce who get great satisfaction from their work (Brisbois & Saunders, 2003; Nickson, Warhust, Dutton, & Hurrell, 2008). NGOs are presented as being socially responsive, flexible, innovative, and better able to identify and respond to community needs than centralised government departments (Dolnicar, Irvine, & Lazarevski, 2008; Edwards & Hulme, 1996; Productivity Commission, 2010). NGOs often enjoy significant goodwill and reputation in the community (Van Slyke & Roch, 2004), allowing them to deliver programs and services that could be problematic for a government agency to provide. Ebrahim (2005a) characterises the advantage of community connection as resource interdependence; that NGOs rely on funding, and funders rely on the reputations of NGOs in order to be able to be effective in delivering the service. O’Shea, Leonard, and Darcy (2002) conceptualise these advantages as social capital
inherent in NGOs through their engagement with and representativeness of the community, and collaboration with other organisations. While some researchers are critical of the assumed superiority of NGO services in light of the limited information available about their internal functioning (Lister, 2003; McDonald, 1999), the acceptance of these advantages has long figured into decision-making about the privatisation of human services (Productivity Commission, 2010).

The increasing tendency for NGOs to deliver human services has come with a shift in the relationship between governments and service providers, and a related increase in compliance processes, evaluation, and accountability. O’Shea (2007) describes a shift from funding human services, to one of purchasing expected service outputs. Baulderstone (2008) suggests that this model of interaction often has a corrosive effect on the relationships between government funders, and the NGOs providing services. The role for government in this arrangement is to set the parameters of service and provide oversight of performance (Cortis, 2007). In line with neo-liberal thinking, this oversight primarily focuses on the economic outputs of human services, emphasising accountability and efficiency (e.g. Auditor-General of Queensland, 2007). This is commonly expressed as results based accountability, which Houlbrook (2011) describes as the use of indicators and performance measures as representative of results in making inferences about the effectiveness of programs. NGOs have found themselves judged purely on economic performance measures, rather than their civic or community mission.

The shift to NGOs becoming the primary provider of human services has brought on significant changes in the sector that potentially undermine or remove the advantages of these types of organisations. O’Shea, Leonard, and Darcy (2007) characterise the effects of competitive tendering and the purchasing of services as a
challenge to the social capital associated with NGOs’ traditional functioning, which is expressed through: the disconnection of NGOs from engagement with the community, the lack of cooperation and collaboration between NGOs; and the loss of collectivism in the organisation. The primacy of neo-liberal thinking in human service NGOs can result in the loss of or dilution of the mission and values of some NGOs (Abramovitz, 2005; McDonald & Marston, 2002), which may challenge the satisfaction service staff in these organisations get from their work (Kirkland, 2001; Wright, et al., 2011). The reduced connection and engagement between NGOs and the community (the loss of social capital) has the effect of reducing knowledge of the local context and ability to identify community needs (Van Gramberg & Bassett, 2005), with needs determined centrally by the government service purchasers (Darcy, Waterford, & McIvor, 2009). Part of the developed social capital of NGOs is the collaboration between organisations, which can take the form of sharing knowledge, but also the sharing of research and evaluation capacity, the opportunity for cross-organisational communities of practice, and the coordination of services to prevent overlap or gaps. Increased administrative and accountability requirements associated with service contracts also reduce the advantage of lower service costs compared to government service provision (Ryan, Newton, & McGregor-Lowndes, 2008). These requirements in turn tend to advantage larger organisations who are able to hire specialist staff (Darcy, et al., 2009; O'Shea, et al., 2007), potentially excluding smaller NGOs that may have strong community representation.

While the direct effects of neo-liberalism on human service provision have been of concern to the sector (Meagher & Healy, 2003), a more subtle and influential force is the effect these changes have on the culture and values of NGOs internally. This concerns the degree to which NGOs have not only complied with the neo-liberal
view of human services, but have actually internalised and taken on these values in
the implicit shared worldview of the organisation.

1.6.2 Neo-institutionalisation as a colonising force
While changes to the relationship between governments and NGOs in the provision
of services have a direct effect, with organisational compliance a requirement of
many funding arrangements, there is a sense that there is change within NGOs
themselves. This change reflects the internalisation of neo-liberal values about the
purpose and functioning of NGOs. This process of internalising shared values and
worldviews within organisations has been described as neo-institutionalism.

Neo-institutional theory (Di Maggio & Powell, 1983) describes how the
implicit values of NGOs can be influenced by organisations they seek legitimacy
from, namely the funders of services. In founding the ‘new’ institutional theory,
DiMaggio and Powell (1983) drew on Weber’s (1952, 1968) imagery of the iron
cage of bureaucracy that perpetuates a homogeneous organisational structure and
culture based on optimal efficiency and effectiveness. However, DiMaggio and
Powell (1983) argued that legitimacy had become the main force for structural
change within organisations. Neo-institutionalism emphasises the importance of a
sense of shared values and worldview within organisations, which Meyer and Rowan
(1991) call institutionalised myth and ceremony. This sociological institutionalism
(Hall & Taylor, 1996) emphasises the socio-cultural construction of meaning that
may guide individual behaviour within institutions. In this tradition Scott (1987)
describes institutionalisation as “the social process by which individuals come to
accept a shared definition of social reality as defining the way things are and/or the
way things are to be done” (p. 496). Institutionalisation requires a level of taken-for-
grantedness of unquestioned realities (Di Maggio & Powell, 1983; Jepperson, 1991),
which are shaped and influenced by discursive practices (Grant et al., 1998).
One of the main themes in the literature is the loss of the distinctiveness and autonomy of NGOs, as they begin to function more like government agencies (Baulderstone, 2008). Through interactions with, and through the need to appear legitimate to the government funders of human services, NGOs are likely to adopt the discourses and structure of the organisations they seek legitimacy from (e.g. Lazarevski, Irvine, & Dolnicar, 2008), and over time institutionalise these implicit values and worldview (McDonald, 2006). However it is important to acknowledge the existence of multiple worldviews or social realities within organisations, McDonald and Reich (2008) explain that “an institutional field can be (and often is) pluralistic in that multiple sub-rationalities can operate within it” (p. 63). In a number of studies, the split between the institutional values of the broader organisation and staff that engage in direct service work with clients are evident (Bowrey, 2007; Chang, 2006; McDonald, 2006). While these direct service staff come from many different disciplines with different values and worldviews (Productivity Commission, 2010), engagement in practice with clients engenders a particular set of values that clashes directly with many of the institutionalised forces of neo-liberalism and managerialism (Dolgoff, Loewenberg, & Harrington, 2009; Fook, 2004; Spence, 2004). Practice represents the major battleground for the clash of values between traditional NGO service delivery and the institutionalisation of the views of social and human services held by the government purchasers of services.

1.6.3 The perceived marginalisation of human service practice
In the human services literature, the change to a purchase model of services (O'Shea, 2007) has been characterised by a broad set of trends centred around the dominance of neo-liberal economic thinking in public discourse and increasingly within NGOs. New Public Management (Boston, Martin, Pallot, & Walsh, 1996) or managerialism is essentially the expression of neo-liberalism, the assumed superiority of private
sector market oriented management practices (Baines, 2004). New public management is associated with an instrumental rationality that is pervasive in the way organisations see practice (Schwandt, 2005). The two main tools associated with this rationality are evidence-based practice and forms of results based accountability. Evidence based practice (Nutley, Walter, & Davies, 2007) reflects a scepticism of professional practice and knowledge, and a preference for scientific evidence to determine practice (Grimshaw, Eccles, Lavis, Hill, & Squires, 2012; Rycroft-Malone et al., 2004). Results based accountability represents efforts to measure the effectiveness of practices in the desired or anticipated outcomes of a program (Friedman, 2004; Rosen, 2003). The enthusiasm for these approaches at high levels of government and management is in stark contrast with the response from the practitioner community (Meagher & Parton, 2004).

Primarily, human service practice has been presented as under attack and oppressed by the dominant neo-liberal outlook (Gray, 2004), with its implicit legitimacy tied to notions of collectivism and social democracy that are currently out of favour (Harris, 2003; Johnson-Abdelmalik, 2011). The value of such work as emotional (Baines, 2011), caring (Meagher & Healy, 2003; Meagher & Parton, 2004), and altruistic (Johnson-Abdelmalik, 2011) labour have been undermined. Singh and Cowden (2009) suggest that the impact of neo-liberalism has been the de-intellectualisation and removal of the critical aspects of human service work, and the representation of such work as a set of skills and competencies. The kinds of critical knowledge thought essential to practice (Baines, 2006; Parton, 2008) have been replaced with characteristics related to efficiency, predictability, and control (Dustin, 2007). The recognition of the complex context of practice and the importance of sound professional discretion has been replaced by a technocratic approach drawn
from the private sector (Ferguson & Lavalette, 2006; Jones, 2005). Human service practitioners with a commitment to ideals of social justice are disdainful about what practice looks like under neo-liberalism. The need for standardised and easily measurable practice has led to a reconstruction of what constitutes legitimate knowledge to inform practice (Harris, 2003), primarily research, evaluation, and performance data.

A number of criticisms have been levelled at attempts to apply a technocratic approach to practice. First, there are false assumptions about the objectivity of data, organisations make value laden decisions about what outcomes are important and how they are measured (Sanderson, 2002a, 2009). The types of values that drive the measurement of practice tend to be top-down (Clarkson, 2010; Jordan & Jordan, 2000) and reflect accountability to funders rather than a broader accountability to the community (Burton & Van Den Broek, 2008; Ebrahim, 2005b). Secondly, converting the complexities of human service delivery into common and meaningful metrics is argued to be reductionistic and superficial (Holland, Thikell, Trepanier, & Earle, 2009). These metrics may also have the effect of favouring programs and practices that narrowly address the measures without necessarily achieving the outcomes they represent. Thirdly, the interaction between measurement and practice is premised on an “instrumental rationality” (Sanderson, 1998, p. 4), a conceptualisation of practice as an object that can only be maintained by impartial analysis (Schwandt, 2005). This view underestimates the capacity of practitioners to create knowledge through reflection or through interpreting research (Arnkil, 2012), or to undertake practice in a way that empowers and acknowledges the self-determination of clients (McCormack, 2004; Titchen & McCormack, 2010). Lastly, given that this measurement rarely has direct implications for practice, it is viewed as
having limited use and relevance to human service providers (Meagher & Healy, 2003).

It is clear that practitioners view evaluation as part of the oncoming storm of neo-liberalism that seeks to undermine professional practice. Research suggests that human service workers take a dim view of evaluation, seeing it as disruptive (Carman & Fredricks, 2008; Taut & Alkin, 2003), geared towards the interests of particular stakeholders (Ebrahim, 2005a), and ultimately irrelevant to practice (Carman & Fredricks, 2008; Meagher & Healy, 2003; Taut & Alkin, 2003). Yet, despite this anxiety and cynicism about the way evaluation is done, there is interest and enthusiasm about evaluation data and its potential usefulness for practice. Practitioners are concerned with their efficacy and engage in their own processes of self-evaluation (Cortis, 2006; Healy & Meagher, 2001). Organisations that have sought to develop the capacity of practitioners to understand and practice evaluative thinking have found success in fostering opportunities for program and practice improvement (Cowley & Good, 2010; Diaz-Puente, Yague, & Afonso, 2008; Whitehall, Hill, & Koehler, 2012). Evaluation researchers have found that anxiety about broad changes in the sector and the role of evaluation in this limits effective engagement in and the use of evaluation amongst practitioners (Donaldson, 2002). The challenge may be in signalling to practitioners the intention to do evaluation that is formative and genuinely participatory.

While practitioners express concern about the loss of autonomy and professional discretion related to these broad trends (Lipsky, 1980; Lipsky & Smith, 1990), some writers have suggested that this represents an emotive overreaction to fairly modest changes (Evans, 2011; Evans & Harris, 2004). They point to the traditional lack of accountability human service professionals and organisations have
had for the effectiveness of their services (Gray, Plath, & Webb, 2009) under the auspices of ‘doing good’ or being socially responsive (Benjamin & Misra, 2006). Particularly in NGOs, there is a history of limited external scrutiny and accountability (Carman, 2010). McDonald (1999) describes evaluation in NGOs as ceremonial and of limited value. In the discourse of evidence based practice, practitioners are characterised as well meaning, but ignorant and ineffectual, while advocates for evidence and evaluation are presented as seeking to turn human services in a mechanical, standardised process. There is a middle ground between these two characterisations through the kind of practice oriented evaluation described by Schwandt (2005). The challenge is for evaluation to be responsive to the interest and needs of practice by pursing a compatible objective, namely social improvement.

1.7 The Challenge for Evaluation

The central problem is the difficulty in evaluating services and programs in a way that is meaningful to organisations and practitioners, and ultimately pursues social betterment. Mark, Henry, and Julnes (2000) have proposed social betterment as the ultimate goal for evaluation, which they define as “improved social conditions, the reduction of social problems, or the alleviation of human distress” (p. 86). While social betterment can be pursued through structural and policy reform, arguably the most direct and powerful means is through direct practice. The challenge is for evaluation to go beyond bureaucratic imperatives such as accountability and efficiency to provide meaningful information for the improvement of services, which Schwandt (2005, p. 102) calls “practice-oriented evaluation” (p. 102).

Schwandt’s (2005) vision for evaluation relating to practice is more like pedagogy than an applied social science; assisting human service practitioners to enhance their ability to deliberate well. He does not object to ‘what works’ type
research but feels this approach has a limited authority on everyday practice. Shaw and Shaw (1997) suggest that the evaluation of services needs to reflect the realities of practice and the tacit knowledge of practitioners, while resting on plausible and falsifiable evidence. More recently, Schwandt (2005) calls for a practice-oriented evaluation based on a substantive rationality that values human outcomes wider than efficiency or effectiveness. These writers seem to be arguing for evaluation that goes beyond accountability to deliver information that is meaningful and reflects the context of practice. The next chapter will examine how evaluation has developed from an applied social science, to a robust and responsive trans-discipline that has increasingly come to focus on fostering learning and improvement and developing the capacity of practitioners to use evaluative thinking for themselves (Diaz-Puente, et al., 2008; Taut, 2007).

1.8 Overview of Thesis Structure

In order to provide the reader with a clear sense of the thesis structure, this section summarises the content of each of the thesis chapters. The first part of chapter one briefly presents the thesis as a whole, setting out the core arguments, the research questions and aims, and the implications. This then turns to providing a sense of the context of human service delivery in Australia, connecting practitioner experiences to the challenges of fostering learning. The shift to NGOs as the primary provider of human services is considered, airing concerns about the institutionalisation of NGOs, and the marginalisation of practice due to broad international trends.

Chapter two deals directly with the development of evaluation from a form of applied social science research to a developed and distinct field of practice (Owen, 2007; Patton, 2008). The current literature emphasises the importance of practicing evaluation in a way that fosters program learning and improvement. This then leads
into a review of *evaluation use*, and the development of *evaluation influence* as a way of understanding and charting the effect an evaluation has had.

Learning from evaluation presents a number of conceptual challenges. Chapter three sets out to define practitioner learning and provide a rationale for why it might be distinct from organisational learning (Argyris, 1999). The literature describes a number of different roles for evaluation in learning, which are outlined in detail. The factors in the literature associated with *evaluation use* and their relevance to the present research are discussed and summarised.

Chapter four outlines the methodology of the research, beginning by establishing the philosophical underpinnings of the thesis. Drawing on moderate social constructionism (Hess, 1997) this thesis employs an abductive research strategy (Blaikie, 2003; Given, 2008) in order to draw out the conditions and barriers to practitioner learning from evaluation. These come from a detailed case study of two evaluations of human service programs, and phenomenological interviews with human service staff. The rationale for the selection of sites and participants is discussed, followed by a description of the data collection and data analysis process.

Chapter five presents the findings in the form of a description and summary of the case studies, leading into the application of the *evaluation influence* framework. Each of the cases are presented using the organisational documents, to provide the reader with an understanding of the organisation, the program, and how the evaluation was conducted. Following this, the analysis of *evaluation influence* in each of the cases is examined, drawing on Mark and Henry’s (2004) *evaluation influence* mechanisms to describe how the evaluation resulted in changes to the organisation and the program. Finally, the experiences and perceptions of the
evaluation are presented in order to develop the participants’ ideas about how each evaluation could have been conducted in a way to foster practitioner learning.

Chapter six provides an analysis of data presented in chapter five, and discusses the conditions and barriers identified in the analysis of the case studies. The chapter suggests that the key factors are related to difficulty in understanding the relationship between evaluation and practice, which underlie issues around practice validity, equitable practitioner engagement in evaluation, acknowledging the critical politics of evaluation, and resolving the clash of inter-organisational priorities in evaluations. This suggests that evaluators can enhance the conditions for practitioner learning by promoting the embedding of evaluation into practice, producing evidence valid to inform practice, making proper use of practitioners as evaluation stakeholders, engaging with the critical politics of organisations and the sector, and advocating for practice relevant evaluation.

Chapter seven concludes the thesis by summarising what was found and the significance of these findings in the context of the literature. This chapter will also discuss some future areas of research extending from the thesis findings.

1.9 Chapter Summary

This chapter summarised and introduced the thesis, followed by providing a broader context for the topic of the research. The chapter acknowledged the importance of evaluation that enhances human service practice, and presents the central thesis question as exploring the conditions and barriers to practitioner learning from evaluation. This question is addressed by undertaking two case studies of evaluations that intended to foster learning, applying Mark and Henry’s (2004) evaluation influence framework to understanding the broad impact of these evaluations, and finally exploring the factors that were important to practitioner learning in the
context of these evaluations. The remainder of the chapter provided a context for the research. Due to an ideological commitment to neo-liberalism, governments are engaged in contracting out human services to NGOs (Australian Bureau of Statistics, 2002, 2009). This shift has resulted in a change of relationship between governments and NGOs (O'Shea, 2007; O'Shea, et al., 2007), a change that brings with it the requirement for performance measurement and evaluation of externally contracted services (Cortis, 2007). The imposition of evaluation tends to be construed by practitioners as part of a broad trend towards NGOs embracing managerialism and bureaucracy (Carman & Fredricks, 2008; Gray, 2004). Despite concern about the effect these trends may have on socially responsive practice, practitioners are interested in understanding their efficacy in terms of human outcomes (Cortis, 2006; Healy & Meagher, 2001). A number of researchers have called for an evaluation that goes beyond accountability to deliver information that is meaningful and relevant to practitioners (Schwandt, 2005; Shaw & Shaw, 1997).
2. CHAPTER 2: PRACTICING EVALUATION FOR LEARNING AND IMPROVEMENT

2.1 Introduction

Contrary to the previous chapter’s depiction of evaluation as a tool of managerialism, contemporary evaluation practice emphasises participatory process and the building of evaluative knowledge and capacity amongst those who engage in the service level work of a program (Hall, Ahn, & Greene, 2012). While in the past evaluation has primarily served as a mechanism for accountability and a source of information for decision-makers, learning and program improvement have become core outcomes for evaluations (Ebrahim, 2005a; Hoole & Patterson, 2008; Torres & Preskill, 2001). Although evaluators have sought to have an impact at a practice level, improvements have been thought about as coming about from the top-down transmission of directives through management structures (Antonacopoulou, 2006; Gould, 2000; Jordan & Jordan, 2000; Sanderson, 2002b). Applying evaluation to the task of improving services directly, at a service practitioner level is lauded as the best way of achieving tangible improvements to the social conditions of service users and communities (Sanderson, 1998; Schwandt, 2005; Shaw & Shaw, 1997).

This chapter will present how evaluators have described and researched the impact of evaluation to provide a foundation to the discussion in chapter three about defining practitioner learning from evaluation. Beginning with defining evaluation, this chapter will then follow into a summary of the study of evaluation impact, moving from evaluation use to evaluation influence. Since the 1970s, evaluation has undergone a transition from being a form of applied research associated with a specific methodology (Guba & Lincoln, 1989), to a trans-discipline that mirrors the
philosophical and methodological diversity of social research (Scriven, 2003).

Following the discussion of this transition, the chapter will then develop the concept of *evaluation influence* as the core framework of this thesis to explore practitioner learning from evaluation. This will include a summary of the studies that have used *evaluation influence* and the implications of these for this research.

### 2.2 Defining Evaluation

The type of evaluation this thesis is concerned with is *program evaluation*. The frameworks and logic of evaluation have developed around the challenge of assessing the value or worth of publicly funded programs (Owen, 2007). While alternate understandings of evaluation exist in other areas of research, and other parts of public discourse, it is generally understood that evaluation is the work of a professional evaluator, who acts much like a researcher in the study of a program (Trochim, 2009b). However, exactly what distinguishes an evaluator’s inquiry from other types of inquiry is vital to a robust definition of evaluation.

In the most general sense, to evaluate is to “find or judge the quality or value in something” ("Evaluate," 2001). Along similar lines, the *Encyclopaedia of Evaluation* defines evaluation as:

> ... an applied inquiry process for collecting and synthesising evidence that culminates in conclusions about the state of affairs, value, merit, worth, significance, or quality of a program, product, person, policy, proposal, or plan. Conclusions made in evaluations encompass both an empirical aspect (that something is the case) and a normative aspect (judgement about the value of something). It is the value feature that distinguishes evaluation from other types of inquiry, such as basic science research, clinical epidemiology, investigative journalism, or public polling (Fournier, 2005, p. 140).

Trochim (2009b) challenges the idea that evaluation must involve judgement (the normative), alternatively suggesting that all evaluations necessitate “… the systematic acquisition and assessment of information to provide useful feedback” (p. 1). A traditional view suggests that an evaluation that does not involve judgement is only a
kind of pre-evaluation (Guba & Lincoln, 1989, p. 29). However, there is a general consensus in contemporary evaluation that these inquiries are forms of evaluation in themselves (Owen, 2007). The proliferation of formative evaluation approaches – often alongside a summative component of an evaluation – show that the improvement or development of a program can be the key outcome of an evaluation (Scriven, 1991; Tessmer, 2008).

*Figure 2.1 Owen’s (2007) Two Definitions of Evaluation. Adapted from Owen (2007, p. 18)*

---

**Definition 1**: Evaluation as the judgement of worth of a program

**Definition 2**: Evaluation as the production of knowledge based on systematic inquiry about a program

This thesis adopts the broader view of evaluation as the production of knowledge. Figure 2.1 presents Owen’s (2007) sense of how the traditional view of evaluation and evaluation as program improvement overlap. Owen (2007) explains that in the broader sense, there are *evaluative inquiries*, that respond to the need for information about a program (p. 17). The acceptance of this definition has led to a broader view of how evaluations can influence change, beyond the production and
dissemination of an authoritative report. Within this definition, is the understanding of evaluation as purely a judgement of worth that is arrived at through the application of a specific evaluative framework (Owen, 2007; definition 1). This logic usually involves establishing criteria of worth, constructing standards, measuring performance and comparing with standards, and synthesising and integrating evidence into a judgement of merit or worth (Fournier, 1995). Evaluation has more or less adopted the broader view, however, in both these definitions evaluation is distinct from research in having to be directly engaged with values.

2.2.1 Values in evaluation
Evaluation is different from research in that it recognises that political considerations cannot be eliminated from the inquiry; both in terms of the construction of the criteria and standards by which a program is judged, and the inferences made based on the information available to the evaluator (Greene, 2000; Markiewicz, 2008). Program evaluation is not objectively technical, nor neutral, it is unavoidably political because it makes assessments about the appropriate use of resources (Faircheallaigh, Wanna, & Weller, 1999) and the worthiness of the goals and outcomes pursued through these resources (House & Howe, 1999). Echoing a series of debates about values in evaluation held in the 1980s (Patton, 1988a, 1988b; Weiss, 1988a, 1988b), Lois-Ellin (2011) summarises the two main approaches to values in evaluation, “a commitment to evaluation as part of deliberative democracy” (p. 279), and “a role for evaluators as sources of unbiased, fair information” (p. 279) in order to serve the public interest. Each of these positions recognises the challenge of values, either by setting out to reconcile and resolve the different perspectives, or by attempting to be value-free.

This thesis takes the view that evaluations of the worth of programs are inevitably political and social constructions, so valid assessments of the merit or
worth of programs should be responsive to community and stakeholder perspectives (Luskin & Ho, 2012). More so than research, the role of an evaluator is to consult and achieve some compromise or consensus on what information will be useful (Patton, 2008; Thomas & Palfrey, 1996). Scriven’s (1991) definition of evaluation reflects the fact that the empirical standards that programs are judged on are constructed:

*The evaluation process normally involves some identification of relevant standards of merit, worth, or value; some investigation of the performance of the evaluands\(^2\) on these standards; and some integration or synthesis of the results to achieve an overall evaluation or set of associated evaluations* (p. 39).

If these standards are not representative of the views and perspectives of stakeholders, evaluations run the risk of being useless (Johnson et al., 2009), or even contributing to the marginalisation of oppressed and vulnerable groups in public discourse (Baur, Abma, & Widdershoven, 2010; Thomas & Palfrey, 1996).

While this view of evaluation is persuasive, there is also a role for evaluations that produce robust evidence. Debates on the nature of credible evidence have been ongoing in the research and evaluation communities, particularly in regards to hierarchies of evidence that hold randomised controlled experiments (or systematic reviews of such studies) as the gold standard (Farrington, Gottfredson, Sherman, & Welsh, 2002; Leigh, 2009; Milewa & Barry, 2005). While the issue is far from settled, particularly across different discipline groups (Donaldson & Christie, 2005; Parker, 2002), there is some recognition that methodological appropriateness should be the guiding principle in the credibility of evaluation evidence (Mark, 2009). This suggests that a legitimate evaluation must be a methodologically reasonable source of evidence in order to make a normative judgement or observation about the issue

---

\(^2\) The term *evaluand* is used within the evaluation community to refer to the subject or object of an evaluation; the program, policy or individuals being valued (Fournier, 1994).
under consideration. What is a reasonable source of evidence will partly depend on the evaluation context and the stakeholders. An evaluator should take into account the types of evidence stakeholders see as legitimate and the classes of evidence compatible with the paradigm or approach of the evaluation. They should also communicate the limitations and weaknesses of particular types of evidence to these groups.

2.3 What is the Intended Use of Evaluation?

The purpose of an evaluation tends to be determined by the funder who sets out the parameters and scope. An evaluator will then make more specific decisions about how to achieve this purpose through the appropriate paradigm (Mertens, 2005), or the approach/form (Owen, 2007; Stufflebeam, 1999). Evaluation may be undertaken in order to meet the expectations of external stakeholders, as an internal effort for organisational learning and continuous improvement, or as an integration of internal and external motivations (Alaimo, 2008). Evaluation often functions as part of a system of accountability between funders and recipients, allowing governments to keep track of the use of funds, the provision of services, and the achievement of desired outcomes (Hoefer, 2000, 2008). Evaluation can also serve as a form of accountability to service users, an often neglected stakeholder of social programs (Baulderstone, 2007; Cortis, 2006, 2007). Evaluation is part of the information available to groups of funders who make decisions about the program, as well as groups within the program who have influence over implementation (Patton, 2008).

The Australian Government Productivity Commission (2010) identifies four main categories of responses regarding the value of measurement and evaluation. Evaluation can improve organisational performance; make information available to funders that can foster greater trust; provide public recognition of the performance of
the sector as a whole; and can produce better policy outcomes by making important performance information available. The relative importance of the different uses of evaluation has been the subject of ongoing research.

In practice, program improvement is an important part of what NGOs use evaluation for. In a survey of 189 charitable organisations in the United States, Carman and Fredricks (2008) examined how evaluation information was used: most frequently it was used to help make changes in existing programs (93%), reporting to the board (82%), and helping to establish program goals and targets (75%). Two-thirds of the organisations reported using evaluation data for strategic purposes, decisions about staffing (68%), to help develop new programs (68%), and to report to funders (67%). Preskill and Caracelli (1997) surveyed 282 members of an American Evaluation Association (AEA) working group on evaluation use and found almost complete agreement with the idea that the purpose of evaluation was to provide information for decision-making (99%) and to improve programs (99%). There were also high levels of agreement with the idea that the purpose of evaluation is to foster organisational learning (88%). A more recent follow up on this study (Fleischer & Christie, 2009) with a much larger sample (N= 1,140) found that members of the AEA thought that evaluations should contribute to organisational learning (84%), help users question assumptions about their practice (73%), increase users’ evaluative thinking (71%), enhance group learning (69%), and enhance individual learning (66%). This later study of the broader membership of the association suggests that evaluators see practitioner learning and evaluation capacity building within programs as an important outcome of their work.
While practicing evaluators acknowledge the importance of program and practitioner learning, the understanding of evaluation being used for more diffuse purposes has developed gradually over time in the research literature.

2.4 How Evaluation Use is Conceptualised in the Literature

While the use of research is important, the success or quality of research does not necessarily depend on its use or impact. However, evaluations that are not used, regardless of their quality, tend to be considered failures (Grob, 2003; Patton, 1997); the job of an evaluator is often not just to produce findings, but to practice in a way that is likely to have an impact (Patton, 2008), and also to go some way towards fostering the implementation of findings (Lawrenz, Gullickson, & Toal, 2007). The development of an understanding of evaluation use has been the subject of significant research and theoretical interest since the 1970s, and has been both responsive to and influential on the modern practice of evaluation. From early decision-based models, conceptualisations of evaluation use have developed to include more subtle impacts and attention to the effect that the process of an evaluation can have. More recently, evaluation influence has been suggested as another way to think about the effect an evaluation can have. This section will help to locate the discussion around evaluation influence, the central framework of this thesis, while acknowledging the place of this concept in the more developed research area of evaluation use.

2.4.1 Use/utilisation: 1970-1986

The impetus for research on evaluation is often attributed to Weiss’s (1972) call for the systematic study of evaluation use in order to improve evaluation practice. Early research tended to conceptualise evaluation use in terms of its direct impact on a decision, in some studies this meant simply examining whether or not
recommendations of the evaluation had been followed (Brickell, 1976; Caplan, 1976; Granville, 1977; Heldt, Braskamp, & Filbeck, 1973). Other researchers explored proxy variables for use such as agreement with evaluation findings (Braskamp, Brown, & Newman, 1978; Brown, Braskamp, & Newman, 1978; Newman, Brown, & Littman, 1979; Thompson, Brown, & Furgason, 1981), perceived usefulness (Alkin, et al., 1974; Glasman, 1984; Lorenzen & Braskamp, 1978; Rossman, Hober, & Ciarlo, 1979), or helpfulness (Goldberg, 1978). Many studies also employed self-report measures for participants to indicate how much impact an evaluation had on their decision-making relative to other influences such as intuition, experience and political beliefs (Dickey, 1980; Florio, Behrmann, & Goltz, 1979; Leviton & Boruch, 1983).

Many of these early studies have been criticised as relying on a flawed standard of evidence, particularly the uncritical use of self-report measures and the lack of triangulation or other means of verifying the behavioural changes attributed to evaluation findings (Leviton, 2003). Reflecting on the definition of use in other studies of the era, Alkin, Daillak, and White (1979) were critical of the short reference points used, arguing that the impact of an evaluation may take months or even years to manifest. Studies also construed recommendations not acted upon as examples of non-use, ignoring that evaluation could inform a decision but not necessarily change it, and that decision-makers can often have good reasons to ignore evaluation findings (Birkeland, Murphy-Graham, & Weiss, 2005). Caplan (1980) was similarly dismissive of the uncritical perspective that utilisation was always a good thing, suggesting that a clear benefit to the use of such evidence needed to exist. Alkin, Daillak and White (1979) were also attentive to the wider impacts or
unintended consequences of an evaluation, and the interactions of people and events that give meaning to evaluation.

While this period of research (1970-1986) is characterised by the conceptualisation of use as a direct and empirically observable phenomenon, other threads of research in this period pursued the idea of evaluation use as a more indirect influence on decision-making. Naturalistic approaches gained some favour during this period (Alkin, 1980), with many studies allowing participants to define evaluation use for themselves (Patton, et al., 1977; Rich, 1979). Early theoretical approaches to evaluation use were also important in reinforcing a broad view of use and driving the shift to a new phase of research. For example, building on principles of communication theory, Newman, Brown, and Braskamp (1980) described a number of stages that preceded but did not necessarily lead to use: attention, understanding, acceptance, and attitude and behaviour change. However, the most important development, which led to new understandings about evaluation use, was the recognition of the complexity of change processes within decision-making bodies.

While not used in the way that evaluators expected, many early studies found evaluation findings to have considerable impact. Despite the apparent disregard decision-makers had for the recommendations of evaluators in their decisions (Alkin, et al., 1974; Caplan, et al., 1975; Patton, et al., 1977), researchers found that such decision-makers valued research and evaluation information highly (Florio, et al., 1979; Weiss & Bucuvalas, 1977; Weiss & Weiss, 1981). Resolving this contradiction, researchers - particularly those engaged in naturalistic or retrospective studies of evaluation use - found that evaluation findings frequently made important contributions to decision-making. First, by influencing the management and practices
within programs (Alkin, et al., 1979; Alkin, et al., 1974; Becker, Kirkhart, & Doss, 1982); secondly, by changing the way problems were understood by decision-makers (Caplan, et al., 1975; Knorr, 1977; Thompson, et al., 1981); and thirdly researchers recognised evaluation as one of many pieces of information that feed into the collective consciousness of decision-makers (Leviton & Boruch, 1983; Patton, et al., 1977; Weiss & Bucuvalas, 1980). Existing definitions were adapted into a typology that reflected the more diverse types of use observed in these studies. There were a number of early variations in terminology (e.g. Turnbull, 1998), but the most enduring taxonomy has been the distinction between instrumental, conceptual, and symbolic/persuasive use.

To an extent, this taxonomy of *evaluation use* borrows from Weiss’s (1979) seven meanings of research utilisation, but tends to be attributed to Leviton and Hughes (1981) or Rich (1977). While the definition of *evaluation use* has developed over time, *instrumental, conceptual, and symbolic use* remain as cornerstones, even in current research and literature. *Instrumental use* began as a direct, documented, and specific use that researchers expected to observe (Rich, 1977, p. 200), but has since come to include the influence the evaluation has over long periods of time and through a variety of indirect agents (Weiss, Murphy-Graham, & Birkeland, 2005). This type of use depends on evaluation results as being the basis of a decision, regardless of the length of time or indirectness of the influence. *Conceptual use* began as the influence evaluation has that cannot be linked to a specific documented use (Rich, 1977). This definition has developed to describe a situation where the use of an evaluation is not direct, but rather the information is absorbed into the common knowledge and comes to form a part of the frame of reference for decision-makers (Weiss & Bucuvalas, 1980). This “enlightenment” (Weiss, 1979, p. 429) has been
characterised as one of the most important means by which evaluation can assert influence (Anderson & Biddle, 1991; Ballart, 1998; Radaelli, 1995). The *symbolic use* of evaluation legitimises decision-makers who have made policies based on intuition, professional experience, or any other of the many influences on policy (Weiss, et al., 2005). Distinct from the previous two types of use, symbolic use does not involve evidence to inform thinking, rather to justify the position already held.

While certainly conceptual use was understood and recognised in this phase of evaluation research, this broader approach to understanding the impact of an evaluation did not become central until after the Cousins and Leithwood (1986) systematic review. This study remains an important marker in the development of evaluation research; Cousins and Leithwood (1986) provide a useful summary of the state of research in their review of sixty-four *evaluation use* studies. Evaluation results were defined as any information associated with the outcomes of the evaluation, including data, interpretations of data, findings, and recommendations, communicated at any point in the research (Cousins & Leithwood, 1986, p. 332). The types of use found in the studies were categorised in the following way:

- Use as Support for Discrete Decisions: Evaluation results had a direct effect on a particular decision. These decisions concerned program funding, the nature or operation of a program, and program management.
- Use as Education for Decision-Makers: Rather than directly affecting decision-making, evaluation results influenced the way decision-makers perceived problems.
- Use as Constituted by Psychological Processing: Use was the degree to which participants thought about or processed evaluation information.
- Potential Use: The dependent variables of some of the studies included represented the potential for use more so than actual use. These studies measured antecedents to use such as agreement or satisfaction with evaluation findings (Cousins & Leithwood, 1986, pp. 341-246).
The studies included in Cousins and Leithwood (1986), while reflecting earlier definitions of evaluation use and focused on results based use, suggest the future development of the influence framework through categories of studies that recognise more subtle effects than direct use. It should be noted however, that the potential users of evaluation as understood by almost all of the studies at the time were decision-makers, rather than human service practitioners. In addition to recognising the importance of a broad definition of use, Cousins and Leithwood (1986) were critical of the limited theoretical base of use and called for the evaluation community to attend to this. The review served to consolidate research knowledge in a way that opened up important issues for the evaluation community to deal with.

2.4.2 Evaluation use/impact: 1986-2000

According to Shulha and Cousins (1997) the most important development following Cousins and Leithwood (1986) was the increasing importance of context in utilisation theory. Evaluation researchers were concerned with how evaluation results interact with other influences in decision-making processes and how evaluators should fit into this process. While contributing to this, Patton (1998) also expanded ideas of use beyond just results and reports, to the experience of participants involved in an evaluation; that change can occur through the process of an evaluation. This observation has in part driven modern approaches to a participatory and collaborative evaluation process. Also during this period, theory concerning the misuse of evaluation developed, distinguishing non-use from misuse (Alkin & Coyle, 1988). These advancements were influential on practice as well as leading to the next phase of research, where researchers sought to theorise and observe the subtle influences of evaluation findings and processes.
One of the most important events in the development of evaluation theory post-1986 was the Weiss-Patton debates. Each being luminaries in the evaluation field, but encapsulating very different visions for the profession, the exchange polarised and energised debate and theory on the ideal nature of evaluation use. Weiss (1988b) argued that aiming for instrumental use ignored the realities of the decision-making context, which tends to be focused on day to day concerns such as the financial solvency of programs, good relationships with staff and funding bodies, and good press for programs. She advocated evaluators accepting the reality of organisational decision-making and focusing on producing sound evaluations, rather than attempting to engage in organisational politics:

*What evaluators should aspire to achieve in the area of utilization is influence, not the status of philosopher-kings whose dictates determine program futures... evaluation is better advised to add understanding to the program, to illuminate the options and likely effects* (Weiss, 1988b, p. 18).

Patton’s (1988a) position was that rather than accepting a diminished role for evaluation in the modern organisation, evaluators should actively engage in fostering the usefulness of evaluation by delivering the information and processes that meet the needs of the commissioners of the evaluation. Patton (1986) had already outlined this position in his book *Utilisation-focused evaluation*, which served as a manual for evaluators engaged in the type of collaborative and consultative processes required for his vision of evaluation.

In retrospect, Shulha and Cousins (1997) link the Weiss-Patton exchange to increased theory and research on the nature of context and evaluation use, particularly in work on evaluation epistemology (Guba & Lincoln, 1989), the political frame of reference in decision-making (Greene, 1990; King, 1988), psychosocial processing of evaluation information (Huberman & Cox, 1990), and organisational culture and learning (Preskill, 1991, 1994; Torres, Preskill, & Piontek,
1996). Examination of the context of evaluation within organisations represents an early linkage between the concept of organisational learning and evaluation.

Equally influential on the theory and practice of evaluation around this era, was the notion of process use being a mode of utilisation; that the experience of participation in an evaluation can be as influential on behaviour as any recommendations or reports (Patton, 1997). Process use has been discussed as a type of use, alongside instrumental and conceptual use (Sandison, 2006). Patton (1997) theorises that process use can lead to four different outcomes:

- improved communications within an organisation due to the illuminating effect of evaluation inquiry;
- the fostering of improvements in the program by way of an intervention that constitutes data collection for the purposes of the evaluation;
- enhancement of the capacity of program participants to be reflective and discriminating in their work; and
- individual and organisational development, leading to the legitimisation of learning as a goal of evaluation.

Process use was particularly influential in the practice of program evaluation, leading to collaborative and participative approaches that aspire to equitable power relationships, organisational learning, and stakeholder empowerment (Cousins, 1996; Cousins & Earl, 1995). Process use remains an important mode of use, reflected in recent models of evaluation influence (Henry & Mark, 2003; Kirkhart, 2000; Mark & Henry, 2004).

The concept of evaluation misuse also received considerable attention over this period, with the recognition of: (a) reasonable or justified non-use, (b) mischievous or deceptive use, (c) the use of poor quality evaluation or misevaluation, and (d) the suppression of quality evaluation for some purpose (Alkin & Coyle, 1988). More recently, Christie and Alkin (1999) followed up with a rethinking of the
concept focusing on the intent of the parties involved and the stage of evaluation the misuse occurs in. Other studies sought to find practical examples of the intentional misuse of evaluation findings; evaluation for purely symbolic reasons (Dial, 1994; Duffy, 1994; Stahler, 1995), the subversion of an evaluation by service staff who resented being subject to evaluation (Vroom, Colombo, & Nahan, 1994), and deliberate non-use of findings (Cousins, 1995; Palumbo, 1994).

Evidence of a rethinking of evaluation in terms of context, the source of influence, and the misuse of evaluation is most salient in the practices of evaluators and the intended uses for which funders and programs employed evaluation. Evaluators were seen to aspire to collaborative and egalitarian relationships with stakeholders (Cousins & Earl, 1995), to foster the skills and qualities of evaluation within organisations (Preskill, 1994), and to actively foster the use of evaluation among intended users (Patton, 1997). Evaluators and the commissioners of evaluations recognised a variety of uses or benefits that evaluation could produce depending on the needs of the program. Through process use and the study of evaluation context, participant learning became an important goal of evaluation, alongside informing decision-making.

2.4.3 Evaluation influence: 2000-present
Around the year 2000, the evaluation community began discussing and debating frameworks of evaluation influence, a possible successor to theories of use. In many ways this approach reflected Weiss’ (1988a) vision for evaluation, while incorporating the changes in practice and understanding that Patton (1997, 1998) had brought to the field. Critical of past definitions of use, a body of theory and research has sprung up in a relatively short time to provide a more nuanced understanding of the influence of evaluation, while acknowledging the importance of the past thirty years of research. Encapsulating existing insights and approaches, evaluation
influence offers a comprehensive framework with which to consider the intended and unintended impacts that evaluation can have, which is particularly important considering the more indirect goals of modern evaluation (e.g. organisational learning, empowerment) and an onus on evaluators to contribute to better social conditions through their work (Henry, 2000).

While evaluators have long expressed dissatisfaction with the definition of evaluation use (e.g. Alkin, et al., 1979; Caplan, 1980), this became the focus of discussion around the turn of the century. Kirkhart (1995, 2000) called the scope and language of use awkward, inadequate, and limiting. Henry (2000) suggests that use has been embraced as the “holy grail of evaluation” (p. 85), which he presents as an unworthy, unhelpful, and self-serving goal that may limit the contribution of evaluation to improving social conditions. The central problem has been the vagueness of the term, with inconsistent definitions existing in the literature (Henry & Mark, 2003; Kirkhart, 2000; Mark, 2004). Evaluation research depends on clear and consistent definitions and language in order to make sense of what is known (Patton, 2000); Henry and Mark (2003) and Kirkhart (2000) argue that the term use tends to underestimate the impact of evaluation by emphasising results based use, and that use suggests an intentionality, immediacy, and directness that may not always exist. Mark and Henry (2004, p. 37) suggest that “contemporary theories of use (or evaluation utilization) are simultaneously impoverished and overgrown”.

Although researchers have attempted to update the concept in line with new understandings about evaluation use, these ad-hoc additional elements lack a clear framework (Christie, 2007; Kirkhart, 2000; Mark & Henry, 2004; Weiss, et al., 2005). Influence has been suggested as a remedy, either as an extension of use (Alkin & Taut, 2003; Kirkhart, 2000) or a replacement for it (Henry & Mark, 2003; Mark &
Henry, 2004). Influence, by contrast, represents an approach to studying the effects of evaluation in the broadest possible terms, across the indirect, unintended, and long-term (Kirkhart, 2000).

By shifting from use to influence, proponents have attempted to legitimise the study of the full impact of evaluation (Kirkhart, 2000). Influence – which according to Henry and Mark (2003) includes “the subset of evaluation consequences that could plausibly lead toward or away from social betterment” (p. 295) – adds to the scope of impact, not to obscure use, but in order to understand the mechanisms and processes that may be antecedents of use. Shifting to a framework of evaluation influence is argued to have the following benefits (Kirkhart, 1995, 2000; Mark, 2003; Mark & Henry, 2004):

- influence provides a definition and a framework that reflects the full impact of evaluation and a cohesive way to organise theoretical and empirical knowledge of the effect evaluation can have on programs;
- By adopting this more comprehensive view, influence allows for the study of implicit mechanisms that affect change, including processes at the individual, interpersonal, and collective levels;
- influence frameworks are oriented around linkages to more developed constructs in other fields of literature such as attitude change, priming, skill acquisition, and persuasion;
- shifting to an influence framework allows for the study of pathways of influence and the study of situations where evaluation failed to affect change; and
- influence is built around social betterment as the ultimate goal of evaluation, rather than use.

Other researchers have argued that influence is better suited to the study of the effects of evaluation in an organisational context (Cousins, 2004; Poth, 2008; Weiss, et al., 2005), drawing on well-established mechanisms of change from the social
sciences (Leviton, 2003; Weiss, et al., 2005). The development of an integrated framework of influence has also been argued as an effective way of acknowledging the broad effects of an evaluation (Caracelli, 2000; Weiss, et al., 2005), in a way that Leviton (2003) suggests is useful for developing evaluation practice. Moreover, this shift is argued to be vital for the evaluation research community, allowing for a more detailed framework that will enhance the evidence base for evaluation practice (Mark & Henry, 2004), which Mark (2001, 2004) has criticised as being overwhelmingly expert-based and susceptible to fads and ideology.

The suggestion of a shift to influence as the conceptual framework with which to study the effects of evaluation has not been without its critics, who argue that replacing use is unhelpful in informing evaluation practice (Alkin & Taut, 2003; Hofstetter & Alkin, 2003; Patton, 2008). Alkin and Taut (2003) suggest that the influence concept is not helpful as it includes events and factors outside the awareness or control of an evaluator. Social betterment, an important part of Mark and Henry’s model (2004) is also criticised as being unrealistic and impractical (Cousins, 2004; Patton, 2008). Hofsetter and Alkin (2003), and Patton (2008) offer program improvement as the purpose of evaluation and the pre-requisite for any social betterment than can follow. The Henry and Mark (2003) conceptual framework of influence has been criticised by McEathron (2008) as a “hodgepodge of unparallel processes” (p. 42) that do nothing to resolve the problems with evaluation use. She concedes that this has been addressed by Mark and Henry’s (2004) framework and the inclusion of levels of analysis (e.g. cognitive/affective, motivational, behavioural), but is critical of the loss of focus on the individual/interpersonal/collective levels, and of a “rational, linear, uni-directional, pro-innovation adoption or use of the evaluation” (McEathron, 2008, p. 43). These
criticisms seem token considering Mark and Henry’s (2004) framework of mechanisms are still organised around the individual/interpersonal/collective level, and that part of analysing the pathways of influence is about attention to more subtle processes that may not result in policy or practice change.

While the relative merits of conceptualising the impact of evaluations in this way remains under debate, *evaluation influence* represents a developed and nuanced framework to build this research around. Despite a long history of research in *evaluation use*, the field of evaluation is no closer to evidence based practice in terms of how to affect social change (Mark & Henry, 2004). *Evaluation influence*, in the form proposed by Mark & Henry (2004), suggests that building a body of evidence begins with recognising the basic mechanisms of influence that can accrete and result in program level change. Through this researchers can develop knowledge about the factors important to different types of influence, and evaluators can adapt their practice to emphasise different types of influence. For example, an evaluator may emphasise individual level attitude change mechanisms if research was available to suggest this approach was more effective with staff resistant to change. The main criticisms of the approach is that it includes factors that can’t be controlled or foreseen by evaluators (Alkin & Taut, 2003), and that the social improvement offered as the driving purpose of evaluation is unrealistic (Hofsetter & Alkin, 2003; Patton, 2008). While certainly some of the factors related to influence may be out of control of evaluators, there is still value in understanding these factors and developing a body of literature that recognises some of the difficult circumstances evaluators practice in. Hofsetter and Alkin’s (2003) objection to social improvement reflects an ongoing conflict in evaluation that connects back to the Weiss-Patton debates discussed in section 2.4.2 (Patton, 1988a; Weiss, 1988b). This thesis takes
the view that evaluation is distinct from social research in being directly accountable to stakeholders for the quality, relevance, and impact. As Henry (2005) suggests social improvement is part of the social compact between evaluators and the community, meaning that evaluators should be responsible for more than just merely if the evaluation improves the functioning of a program.

2.5 The Evaluation Influence Frameworks

Implicit in the change to influence are the conceptual frameworks that provide some substance to what would otherwise be a point of semantics. Figure 2.2 presents these frameworks in order to contrast the difference between Kirkhart (2000), Alkin and Taut (2003), Henry and Mark (2003), and Mark and Henry (2004). Consistent with their argument for influence to be understood as an extension of use, Kirkhart (2000) and Alkin and Taut’s (2003) models incorporate use, with awareness as the threshold for influence. Henry and Mark (2003) and Mark and Henry’s (2004) frameworks construct influence as a series of mechanisms that cut across understandings of use.
Figure 2.2 A Comparison of Evaluation Influence Frameworks (adapted from Alkin & Taut, 2003; Henry & Mark, 2003; Kirkhart, 2000; Mark & Henry, 2004)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Intention</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unintended</td>
<td>Intended</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aware Unintended</td>
<td>Aware Intended</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Source</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Process</td>
<td>Result</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Time</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Immediate</td>
<td>End-Of-Cycle</td>
<td>Immediate</td>
<td>End-Of-Cycle</td>
</tr>
<tr>
<td>Long-Term</td>
<td>Long-Term</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Level of Analysis</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Individual</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group</td>
<td>Individual</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Interpersonal</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>General Influence</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cognitive and Affective</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Motivational</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Behavioural</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Kirkhart’s (2000) framework suggests three dimensions: intention is simply if the evaluator intended the influence; source reflects the idea that the process and the findings of an evaluation can have influence; and time which is split into immediate, when the evaluation is concluded, or longer term. Alkin and Taut’s (2003) framework slightly differs with the inclusion of awareness, which they suggest represents the difference between use and influence.

Over two papers, Henry and Mark (2003), and Mark and Henry (2004) have refined a more sophisticated framework, building on Kirkhart’s (2000) work, but emphasising the mechanisms through which evaluation has an influence and the
outcomes of such influence. Henry and Mark (2003) present influence as interconnected change mechanisms at the individual, interpersonal, and collective levels. Mark and Henry (2004) elaborate on this model categorising influence into families of similar mechanisms (general, cognitive & affective, motivational, behavioural)\(^3\). A key idea in these frameworks is the interconnection of influence as a kind of chain reaction of events. This framework aims for a comprehensive understanding of the complex, contextual, and often convoluted series of processes that can lead to change and just as often lead to no observable change.

An important feature of Mark and Henry’s (2004) framework is the linkage to more established constructs in research and theory in psychology, political science, organisational behaviour and sociology. Drawing on knowledge and terminology from other disciplines allows for the enrichment and integration of evaluation research with parallel areas of inquiry. Mark and Henry (2003) also argue that the influence framework will potentially serve as a stimulus for systematic inquiry to inform evaluation practice as an antidote to “expert-based... evaluation practice” (Mark, 2004, p. 1). They suggest that evaluation influence represents an opportunity to develop a structure of issues for research involving simple questions that can be built into a collective body of knowledge.

### 2.6 Evaluation Influence Research

The development of a body of evaluation influence research has been slow, indeed Murphy (2007) and Poth (2008) remarked on the lack of empirical support and the lack of take-up in research. Since then a small body of literature has developed around exploring evaluation influence in a variety of settings. For many of the studies, influence has been construed as an attention to the broader impacts of an

---

\(^3\) These will be discussed in more detail in chapter four.
evaluation without drawing on any of the frameworks discussed above (Alexander, 2003; Allen, 2010; Cooksy & Caracelli, 2005; Frey, 2010; McEathron, 2008; Morabito, 2002; Poth, 2008; Vanlandingham, 2011; Weets, 2008). Other studies have examined *evaluation influence* through case studies, surveys, or citation analysis while drawing on the frameworks. These are discussed in more detail below.

### 2.6.1 Case studies

The case studies included have employed *evaluation influence* in an attempt to discover the broader effects of evaluation in the context of a program. While some studies drew on *evaluation influence* more substantially, others merely acknowledged the concept in the literature and used the definition to broaden the scope of the effects they were looking for in their case studies. Bamberger (2004) examined *evaluation influence* in eight development programs by the World Bank through an attribution analysis. Drawing on surveys, interviews, and document analysis, he explored how each evaluation influenced the way policy was implemented. Frey and Widmer (2011) developed a scale of influence in order to measure the impact of systematic evidence in revisions of Swiss government policy. In a brief and informal analysis of three evaluations, Henry (2003) illustrates the kinds of situations where evaluation may be influential.

Despite lacking the conceptual depth of the other frameworks, a number of case studies have used Kirkhart’s (2000) framework in analysing influence. Benjamin and Misra (2006) undertook thirteen interviews with staff within agencies that funded non-profit services to examine the influence of performance accountability. The researchers prioritised the investigation of changes in service practice resulting from *evaluation influence*, as well as second order changes in discussion and discourses. Rebolloso, Baltasar, and Canton (2005) investigated the
influence of two public education evaluations, comparing a capacity building approach to a more traditional evaluation design, and categorising their observations.

A number of studies used the earlier Henry and Mark (2003) framework, which provides similar mechanisms to Mark and Henry (2004) as shown in Figure 2.2, but without categories of change processes (e.g. cognitive, motivational, behavioural). Cowley and Good (2010) applied the Henry and Mark (2003) framework to examine the influence of evaluation on education staff attitudes and behaviour related to their technical assistance work. Diaz-Puente, Yague, and Afonso (2008) and Diaz-Puente, Montero, and Carmenado (2009) examined how influence played out over a ten year period in a series of evaluation capacity building projects in rural Spain. Changes in attitudes and actions at the individual, interpersonal, and collective levels were discussed, along with some description of change mechanisms (e.g. capacity building, salience) linked to the evaluation findings and process. These studies also went to some effort to study the pathways of influence; tracing the interaction between different mechanisms. Incorporating Valovirta’s (2002) observations about argumentation into the interpersonal level of influence, Lehtonen (2010) built on Henry and Mark’s (2003) framework to explore influence through interviews, a document analysis and a stakeholder workshop. Oliver (2008) undertook a study of the evaluation of international NGOs’ emergency responses, developing an in-depth criteria of mechanisms from the framework described by Henry and Mark (2003). The researcher used this criteria to identify influence mechanisms and their connection through an analysis of evaluation reports and interviews with key stakeholders (Oliver, 2008).

Using Mark and Henry’s (2004) later framework, Cheng (2006) undertook a case study of the influence of two evaluations of literacy programs. She discussed the
influence mechanisms observed in the research and reported some success in tracing some of the pathways of influence that followed from the evaluation. Fjellstrom (2007) used the same framework in describing the influence of a collaborative evaluation of a teaching initiative, highlighting some of the pathways of influence observed. Finally, Weiss et al. (2005) undertook a long term case study of the influence of evaluations on the high profile Drug Abuse Resistance Education program in the United States, drawing on both use and influence concepts. Influence was deployed specifically in order to attempt to trace some of the pathways of influence observed across sixteen case studies. Weiss et al. (2005) reported this as being challenging due to the difficulty in connecting events.

2.6.2 Surveys
Burr (2009) drew on evaluation use and influence in developing a survey examining the effect that evaluation had on project directors of seventeen university preparation programs. The survey items reflect the levels of influence and the change mechanisms of Henry and Mark (2003). Swanson & Barlage (2006) drew on evaluation influence in attempting to identify the most influential studies in education policy through a survey of subject experts. However the meaning of the term influence was left to the participants to define in their responding.

In a departure from other approaches to studying influence, Christie (2007) undertook a simulation study of how different types of evaluation data influence decision-makers at the individual level. This simulation involved nine scenarios with different forms of evaluation evidence and the participants’ survey responses on how influential the data were, taking into account the participants’ pre-existing beliefs in the efficacy of the program described.
2.6.3 Citation analysis
Greenseid, Johnson, and Lawrenz (2008) and Greenseid (2008) used citation analysis to assess the influence of a number of education evaluations. In both studies influence is measured as the citation of particular evaluations in their related fields of study, which excludes process influence (Alkin & Taut, 2003; Kirkhart, 2000). Swanson & Barlage (2006) also undertook a citation analysis of studies identified by their sample as being highly influential, which were included in their three dimensional index of influence.

2.7 Evaluation Influence as a Research Approach
While there is a developing literature on evaluation influence, Mark (2004) suggests that in order to address core questions of the effects of evaluation practice, researchers should begin with simple research questions to build a body of evidence. Largely this has not occurred, with much of the literature constituted by conceptually broad case studies across many different areas of study, with much variation in methods, definition and procedures. While many researchers have employed influence ideas, this has not occurred in the context of a clear direction for research with implications for evaluation practice. While the use of case studies is not problematic in itself, and indeed influence seems to be a concept well suited to retrospective case studies, research on evaluation influence has a number of significant limitations:

- Much of the research is built on the investigation of influence by the same individuals that conducted the evaluation, often based primarily on their experience of undertaking the evaluation (e.g. Diaz-Puente, et al., 2009; Diaz-Puente, et al., 2008; Weiss, et al., 2005). This is clearly problematic, with evaluators having a stake in presenting their work as being influential;
- The methodological rigour of these studies varies greatly. Some provided only limited description of method, and no clear operational definitions of
influence or change mechanisms (Henry & Mark, 2003). Although some studies reported using case study protocols, there is a need for clear and explicit reporting of method;

- Many studies rely primarily on self-report by organisational stakeholders who may have a direct interest in presenting a narrative of an organisation that is receptive to evaluation evidence. While almost all the studies cited used organisational documents along with interviews, there is a need to address directly the desirability of being an evidence-based organisation in studying influence; and

- Some awareness of how the timeline of the research may affect the findings may provide some perspective for findings. Studies that allowed a longer interval following an evaluation seem to be more likely to observe the link between these individual changes and collective change (Diaz-Puente, et al., 2009; Diaz-Puente, et al., 2008; Oliver, 2008). Weiss et al. (2005) suggest that in their study the interval had been too long (2-8 years after the events) to adequately capture individual and interpersonal change.

Not all of the studies reviewed set out to examine the relationship between the influences they observed, those that did reported varying degrees of success in following the interaction and interrelationship between different mechanisms. While finding some examples of clear threads of influence, Cheng (2006) and Weiss et al. (2005) reported the exercise challenging due to the intertangling of the threads of influence. Five studies included in the review were more successful in tracing influence, with each beginning their research fairly soon after the evaluation began (Diaz-Puente, et al., 2009; Diaz-Puente, et al., 2008; Fjellstrom, 2007; Lehtonen, 2010; Oliver, 2008). This may suggest that researchers should investigate influence alongside or closely following an evaluation in order to capture the individual level mechanisms that tend to begin longer chains of influence.

*Evaluation influence* has been an important development in the past decade of research on the impact of evaluation. Building on long-standing dissatisfaction with
the definitions of use in theory and the research literature, the proponents of this change have made a case for understanding the effects evaluation can have in the broadest sense, in order to enable evaluation researchers to better describe and understand what occurs during and following an evaluation. *Evaluation influence* represents a new and developing approach to understanding the impact an evaluation has. Given the focus of the present research is how evaluation can foster learning at a practice level, the kinds of mechanisms contained within the *evaluation influence* frameworks (Henry & Mark, 2003; Mark & Henry, 2004), make it ideal as a conceptual foundation for this thesis.

### 2.8 Chapter Summary

This chapter outlined the development of evaluation towards the acceptance of a collaborative and participative model of practice, and described the development of the *evaluation influence* literature. This discussion has focused on how influence has been used in research in order to inform the definition and conceptualisation of influence in this research. Discussion of the findings of these studies and their implications for practitioner learning will be explored in the following chapter (Section 3.3).

This thesis takes the view that evaluation is constituted by the production of useful knowledge from some form of systematic inquiry (Owen, 2007), which is consistent with the way most evaluations are used (Carman & Fredricks, 2008; Fleischer & Christie, 2009; Preskill & Caracelli, 1997). Evaluations are described as being political and social constructions, meaning that valid assessments of the merit or worth of programs need to be responsive to community and stakeholder perspectives (Luskin & Ho, 2012). The chapter described the development of the *evaluation use* literature, from early stages that conceptualised use as a discrete
decision (Caplan, 1976), to the recognition of conceptual and symbolic use (Leviton & Hughes, 1981; Rich, 1977). This then led to increased research on theory in evaluation (Shulha & Cousins, 1997), recognition of the importance of process use (Patton, 1998), evaluation misuse (Christie & Alkin, 1999), and participatory approaches (Cousins & Earl, 1995; Patton, 1997; Preskill, 1994). The development of *evaluation influence* as an alternative to use was then discussed, emphasising the work of Mark and Henry (2004) and Henry and Mark (2003) and their *evaluation influence* frameworks. The chapter concludes with a review of studies that have employed the *evaluation influence* frameworks.
3. CHAPTER 3: LEARNING FROM EVALUATION IN THE NGO AND HUMAN SERVICE CONTEXT

3.1 Introduction

This thesis takes the position that practitioners can and should be the focus of efforts to promote learning from evaluation. Chapter two described the development of evaluation towards recognising the importance of program and practice improvement. How these improvements play out at the practice level can result in changes that impact directly on the experiences of service users. The concept of evaluation influence provides a clear framework with which to understand the effects of evaluation at a practice level. This thesis focuses on a specific approach to these effects, practitioner learning from evaluation; that is, how evaluation fits into practitioners’ processes of reflecting on their own practice.

The chapter will begin by building a definition of learning from evaluation. Practitioner learning has received limited conceptual development in the evaluation literature, although there are some well-established links to the organisational learning literature (Owen & Lambert, 1995; Preskill, 1994; Torres & Preskill, 2001). Evaluation operates as a trans-discipline, working across and providing tools to other disciplines (Scriven, 2003), therefore the definition of learning will inevitably cross disciplinary boundaries. Learning in the context of this thesis refers to changes in practice that ideally translate into improvements either in the program’s performance or in the broader social conditions of service users. These changes come from knowledge acquired through an individual’s reflective process (Antonacopoulou, 2001) which can be prompted by information identified directly (individual level
learning), through the interaction of individuals (interpersonal level learning) or by the organisation (collective level learning).

Following on from defining learning, the chapter will then turn to examine how evaluation researchers have described the role of evaluation in learning. First, evaluation can serve a traditional role providing outside advice about the effects of particular actions (Owen & Lambert, 1995). This role emphasises organisational learning and the dissemination of knowledge through learning systems (Buckmaster, 1999). Secondly, evaluation can be the catalyst for examining the implicit assumptions that underlie daily routines (Benjamin & Misra, 2006; Preskill, 1994; Shaw & Shaw, 1997), which provides opportunities for learning embedded in practice. Thirdly, developing evaluation capacity (Naccarella et al., 2007) amongst practitioners enhances their ability to use evaluation information and evaluative thinking as part of their daily practice and reflection. Finally, evaluation can empower and legitimise marginalised perspectives by recording and disseminating the experiential learning of human service practitioners (Schwandt, 2005). This provides essential information exchange between groups involved with the implementation of the service and the broader organisation (Hoole & Patterson, 2008).

With a sense of how evaluation fits into processes of learning, the chapter will then explore the factors associated with practitioner learning from evaluation. While the evaluation use literature has a long and extensive history, practice level learning remains relatively unexplored. This section draws on existing bodies of research exploring the conditions associated with the use and influence of evaluation, in order to inform the analysis for the present research. These factors tend to be categorised in terms of evaluation implementation and decision-making context.
(Cousins & Leithwood, 1986), but more recent studies have introduced categories of stakeholder involvement (Johnson, et al., 2009), and more detailed sub categories acknowledging individual level factors (Karan, 2009), conditions for organisational learning (Botcheva, White, & Huffman, 2002; Murphy, 2007), and leadership (Alaimo & Van Slyke, 2006; Murphy, 2007; Preskill, Zuckerman, & Matthews, 2003).

### 3.2 Defining Learning in the Context of Evaluation

This thesis concerns learning, but in relation to what it means conceptually to learn from an evaluation. This presents a challenge considering evaluation is described as a trans-discipline (Scriven, 2003), meaning that evaluation works across disciplines that have different understandings of learning (Argyris & Schön, 1978; Fleming, 2011; Schon, 1983). The evaluation literature describes learning broadly, but primarily in terms of organisational processes, which is unsurprising considering evaluations are for the most part intended to effect collective change (Owen, 2007). Individual level learning is discussed primarily as a necessary pre-cursor to organisational learning (Levitt & March, 1988), however both levels require the individual to reflect on the information in order for learning to occur (Antonacopoulou, 2001).

Learning in the human services context ideally translates into the capacity to improve practice. In one sense, improvement means simply the increased effectiveness or efficiency of the program towards its stated outcomes. However, there is a broader picture of improvement to consider, of “…social conditions, the reduction of social problems, or the alleviation of human distress” (Mark, et al., 2000, p. 86), where ideally the desired outcomes of a program align with broad social
improvement. For this thesis, the product of learning from evaluation includes both programmatic improvements and social improvements.

Emphasising practice level change, improvements come about through changes to the behaviour of the practitioner, which can be prompted by: (a) individual level learning; (b) group level learning; or (c) organisation level learning that has produced a change by disseminating knowledge to individuals, or through organisational mechanisms (e.g. policy, practice guidelines).

3.2.1 Individual learning

Individual learning emphasises the agency of the individual in processing and interpreting information, and the importance of their subjectivity in this process (Palincsar, 1998). While the individual is inevitably part of an organisation and a work group, and will share values with these groups (Hwang & Colyvas, 2011), the level of individual variation in a lifetime of experiences means that people do not understand things in the same way (Berger & Luckman, 2011). While the onus for individual learning may come from external sources (e.g. evaluation data), and this process may be influenced by organisational values, this thesis conceptualises practitioner learning as an individual process.

Much like collective learning, individuals learn within a specific cultural context. Hedberg (1981) draws on Hegel to explain that individuals and groups create a perceptual frame that influences the recognition of problems, thinking around how to resolve them, and the learning that results. The relationship between these perceptual frames and learning can be dialectical; learning occurs within the frames, yet also changes the frame in the process (Gherardi, 2006). That said, Antonacopoulou (2006) suggests that organisational context can frame learning within a dominant structure, creating an understanding of learning that does not question implicit or taken for granted norms. This links to theory around neo-
institutionalisation (Di Maggio & Powell, 1983), where collective norms and values can restrict or control an individual’s thoughts or behaviour (Frumkin & Galaskiewicz, 2004; Meyer & Rowan, 1991). In short, even learning that takes place at an individual level occurs within an organisational context with norms that may influence the nature of the learning.

As mentioned, the evaluation literature has well developed connections with organisational learning. In the context of organisational learning, individuals are the active agents of learning; they identify problems, solve these problems through their understanding, implement their solution, and evaluate the effectiveness of their actions before beginning the cycle again (Watkins, Marsick, & Johnson, 2003). Individual learning as a component of organisational learning primarily concerns experiential or constructivist learning by workers “... where the learner reflects on lived experience and then interprets and generalises this experience to form mental structures” (Fenwick, 2001, p. 248). This interaction of action and reflection serves to challenge assumptions and values (Schwandt, 2005), resolve doubt or ambiguity (Argyris & Schön, 1996), and ultimately create new knowledge and practices encoded from experience.

3.2.2 Interpersonal learning
Senge (1990) and Argyris and Schön (1996) make a distinction between individual (as discussed above), group, and collective learning. In groups, discussion and exchanges can occur that challenge individual knowledge and draw out its assumptions, values and beliefs. The social process of engaging workers in understanding and reflecting on the meaning of their work is often discussed in terms of communities of practice (Wenger, 1998), where interaction and exchange results in meaningful learning. This learning is situated in a socio-cultural context where the
identity and connection of the individual to the community is central (Gherardi, 2006).

### 3.2.3 Collective learning

The premise that a collective can learn has not gone unquestioned; Kim (1993) highlights the risk of anthropomorphising organisations. However, collective level learning centres around the relatively uncontroversial idea of individuals learning and performing learning functions for the organisation (Argyris & Schön, 1978). Organisational learning requires the processing or encoding of information into behavioural change (Levitt & March, 1988), or the use of knowledge to improve practices through collective mechanisms (Fiol & Lyles, 1985). This type of learning is embedded within the shared meaning and cultural practices of people within organisations (Crossan, Lane, White, & Djurfeldt, 1995). Despite the organisation legitimising and distributing particular types of information, learning depends on the individual and their personal process of reflection (Antonacopoulou, 2001).

Organisational level learning differs from individual and group learning, as it is dependent on the sharing and dissemination of knowledge between individuals as part of a learning system (Daft & Weick, 1984). Through the accumulation of knowledge in documents, policy, practice guidelines, strategic plans, and procedures organisations develop a shared mental model or understanding that forms part of the organisational memory (Schulz, 2001; Weick & Roberts, 1993). Like individual learning, organisational learning occurs as a social process, with structures, systems, and other contextual factors impacting on how information is shared and valued (Hedberg, 1981; Pawlowsky, 2001). Torres and Preskill (2001) emphasise the importance of learning being integrated with work activities, and the culture and systems within the organisation, invoking an alignment of values and attitudes amongst individuals.
Argyris and Schön’s (1978) description of single and double loop learning have been formative in organisational learning. Learning in this theory involves a change to routines and practices (which constitute organisational knowledge) and their underlying assumptions. Single loop learning is described as a feedback control system where learning is constituted by detecting and correcting errors in practice knowledge (Argyris & Schön, 1978). This type of learning is concerned with improving effectiveness within the scope of existing norms and assumptions. When the examination of underlying assumptions is part of the inquiry process, Argyris and Schön (1978) term this double loop learning. Beyond correcting errors, the organisation modifies the cultural norms of the organisation to produce a change to the values or perceptual frame of the organisation (Argyris & Schön, 1996; Levitt & March, 1988). These implicit understandings can exist at both the individual and organisational level (Argyris & Schön, 1996); however changing values and beliefs at the individual level can be particularly challenging (Schein, 1992), particularly where the change is contrary to a person’s worldview (Martin & Clark, 1990).

Argyris and Schön (1996) also describe deutero learning as when organisations undertake both single and double loop learning in order to change the environment and processes of learning. Recently, triple loop learning was suggested by Ameli and Kayes (2011) to describe changes occurring from the interaction between two partnered organisations; that an organisational partner can contribute to learning processes through practices, structure, and culture.

the information comes to practitioners through their own discovery or experiences, through group discussion and reflection, or through organisational information systems, ultimately it is up to the individual to process and ‘digest’ this information and make the resulting changes to behaviour. Attempts to disseminate information to practitioners for their learning may amount to providing ‘pre-digested’ information; however, evaluation has a number of different roles to play in learning across the different levels of organisations.

### 3.2.4 Learning from evaluation across organisational levels

Learning takes place in a social context (Easterby-Smith, Thorpe, & Lowe, 1991), with meaning derived from shared understandings and implicit theories about the world (Levitt & March, 1988). Evaluation seems to play a number of key roles in engaging with the social context of learning: providing outsider advice (Buckmaster, 1999; Ebrahim, 2005b; Owen & Lambert, 1995; Schwandt, 2005), the examination of implicit assumptions (Owen & Lambert, 1995; Preskill, 1994; Shaw & Shaw, 1997), the promotion of learning amongst practitioners (Schwandt, 2005; Shaw & Shaw, 1997; Torres & Preskill, 2001), and the empowerment of marginalised groups (Ebrahim, 2005b; Hoole & Patterson, 2008).

Traditional approaches to evaluation emphasise the provision of scientific information to organisations, helping them to understand the effects of their actions (Owen & Lambert, 1995). Schwandt (2005) describes this type of evaluation as applied social science that sets out to determine if practices bring about their desired effects. This approach is presented as objective, an assumption that may not be shared by different stakeholders (Cortis, 2007; Ebrahim, 2005b; Taut & Alkin, 2003). The dissemination and implementation of evaluation in this approach flow into planning systems from the top down, so that learning from evaluation is written into the organisational memory (Buckmaster, 1999) and put into practice by
individuals. However, this supposes that all organisations have well developed systems to retain organisational learning (Ebrahim, 2005b). Even when these systems exist, connecting learning to the practitioner level is difficult. Such knowledge is unlikely to be used unless individuals can see the improvements in their work as a result and the potential benefits for service users (Edwards & Hulme, 1996). This approach also fails to attend to the internal politics of organisations and the broader political context of practice. By attempting to stay neutral to the politics and conflicting perspectives within an organisation, this type of evaluation can risk lacking usefulness and relevance (Patton, 1997), and can further marginalise groups within the organisation (Abma, 2006; Meagher & Healy, 2003). Evaluation approaches that are not attentive to organisational context and seek to foster change through disseminating results to a small number of individuals at the top of an organisation will likely foster limited learning (Johnson, et al., 2009).

Another role for evaluation in fostering learning is as the force to examine and reflect on implicit assumptions embedded in daily practices. Program clarification has become an important part of evaluation, making explicit the values and assumptions that underlie routines (Owen & Lambert, 1995). Formative activities such as developing a program logic have become increasingly valued in evaluation practice as a precursor to impact evaluations (Owen, 2007). Promoting reflection on the implicit ways of doing things in an organisation can be part of the activity of an evaluation. Preskill (1994) suggests that evaluation is inherently a tool for double loop learning, providing the impetus and energy of bringing people together to reflect on practices and question accepted wisdom. Similarly Shaw and Shaw (1997) argue that evaluation can serve as a way to encourage the questioning of implicit assumptions by providing a space to “... talk about what is seldom spoken
about” (Altheide & Johnson, 1994, p. 493). Benjamin and Misra (2006) cite the example of evaluations in NGOs providing a chance to get past the idea of ‘doing good’, to promote critical discussions that can lead to constructive feedback. By making individuals’ and groups’ understandings and assumptions clear, learning can result from a dialogue that clarifies and tests their understanding of the world (Guba, 1990). This suggests a kind of constructivist learning (Mayer, 1999) that comes from the sharing and discussion of how individuals, groups, and organisations see their work, and what they think is important. Considering the socio-cultural context of learning in organisations this role is particularly important, as it may play a part in resolving conflicting perspectives and obtaining a sense of the theories in-use by workers (Argyris & Schön, 1978, 1996).

Evaluation also has a role in learning through the promotion of evaluative thinking amongst individuals and groups in the organisation. Shaw and Shaw (1997) argue for a bottom up form of evaluation to inform social work practice through evaluative thinking and reflective practice. Evaluation can play a part in enhancing tacit knowledge and otherwise building capacity for practitioners to undertake, understand, and respond to evaluation (Stevenson, Florin, Mills, & Andrade, 2002). Evaluation capacity building includes the capacity to conduct evaluation, and to participate in evaluation more effectively (Naccarella, et al., 2007). Considering a broader view, Schwandt (2005) describes evaluation like pedagogy, where it “... helps practitioners understand the kinds of evaluation decisions they face and enhance their ability to deliberate well” (p. 99). This role for evaluation is grounded in practical action, and acknowledges the complex interrelationships of people and organisations (Forester, 1999), and can help individuals gain an appreciation of the interdependence of their work and knowledge (Preskill, 1994). While it may be
managed centrally in specialist evaluation units, the capacity to do evaluation is
divested, allowing the practice of evaluative thinking to be integrated into daily work
(McDonald, Rogers, & Kefford, 2003). Forester (1999) suggests that all professional
undertakings bring an imperative to think evaluatively, to critically question values,
practices and goals.

Evaluation has an important role in not only the sharing and discussion of
different perspectives (Guba, 1990), but also in empowering some perspectives that
might otherwise be marginalised in the organisation. Human service practitioners in
particular report being marginalised in modern service organisations (Meagher &
Healy, 2003; Meagher & Parton, 2004), particularly in the context of the perceived
primacy of managerialist thinking (Harlow, 2003; Tsui & Cheung, 2004). Evaluation
has a role in fostering learning through engaging service level staff who possess the
most direct experience with the core work of the organisation (Ebrahim, 2005b).
Staff that work directly with clients in the front line of the program are potentially
the richest source of knowledge about practice; without feedback through the sharing
of knowledge from these staff, the organisation essentially operates blindly. Some
shared understanding and information exchange between leadership and practice is
essential for all other types of learning to be layered on (Edwards, 2002). Evaluation
should contribute to a common understanding of the people and issues the program is
concerned with or risk irrelevancy (Hoole & Patterson, 2008). This role for
evaluation centres around the sharing and legitimisation of practitioner perspectives
as a form of knowledge, and drawing on this to foster learning throughout the
organisation.

Preskill (1994) also suggests a number of other roles for evaluation in
organisational learning. Depending on how it is undertaken, evaluation can signal
organisational commitment to a culture of evaluation and reflection. Evaluation can itself serve as part of the organisational memory that stores knowledge and information (Preskill, 1994). Evaluation that creates opportunities for participation and knowledge sharing also increases interest in the findings, enhancing understanding of the evaluation as well as fostering the influence of findings (Torres & Preskill, 2001).

3.3 Factors Associated with Practitioner Learning from Evaluation in the Literature

With a sense of how evaluation is understood to fit into processes of learning, this section will begin to unpack the factors that seem conducive for practice level learning to take place. While the evaluation research literature has long sought to understand the factors associated with the use of evaluation, there is little compelling evidence of the relative importance of the many factors explored (Leviton, 2003). Mark and Henry (2004) suggest that in part, the failure to build a body of evidence to inform evaluation practice has resulted from problems with the definition of evaluation use. Mark (2004) challenges researchers to develop an evidence base for evaluation practice, drawing on the distinct mechanisms of change organised under Mark and Henry’s (2004) evaluation influence framework. This endeavour is relatively new, yet there is a developing body of research drawing on influence concepts (Herbert, 2011). As pointed out by Alkin and Taut (2003) it is not a desirable state of affairs to abandon existing knowledge in order to adopt this new framework and conceptualisation. Herbert (2011) observes, there is work to be done in integrating the use literature to inform evaluation influence.

This thesis is novel as it focuses on a different agent in the organisation - human service practitioners - and is oriented towards understanding the influence of
evaluation through reflective processes that lead to learning. While there is a wider challenge in integrating the evaluation use literature with evaluation influence, for this thesis there is some difficulty in reconciling existing research about the factors associated with use/influence with practitioner learning from evaluation. Much of the research has examined the use/influence of evaluation from the perspective of decision-making, which is how evaluation can affect the attitudes and behaviours of senior people in a decision-making role. While some analogies can be drawn between this role, and the decisions that social workers, case workers, psychologists etc. make on a daily basis, fundamentally practice and organisational decision-making represent two very different phenomena. In reviewing the literature, the aim is to identify factors that might be of relevance for the analysis of this research, taking into account the definition of use/influence used in the study. The degree to which research involving evaluation use can be generalised to different understandings of effecting change is not yet clear (Frey & Ledermann, 2010). This section unpacks the factors associated with the use and influence of evaluation in order to highlight some factors that may be of importance for practitioner learning from evaluation.

The following sections will summarise the literature on what factors seem to affect the impact of evaluation, separated into the bodies of literature on evaluation influence, and evaluation use. The evaluation influence literature is still developing however some notable studies have sought to explore the important factors for particular types of influence. The presentation, timing and availability of evaluation reports were emphasised in Oliver’s (2008) study of emergency response INGOs, Alexander (2003) also found that factors related to the conduct of the evaluation were most important. Empowerment and engagement were highlighted as important
in changing attitudes about evaluation in Diaz-Puente, Yague, and Afonso (2008), and Diaz-Puente, Montero, and Carmenado (2009). Christie’s (2007) simulation study found all types of evaluation data affected the decisions of individuals, the discipline background and existing attitudes about the program affected the type of data most likely to influence participants. The much more extensive body of literature on evaluation use begins with a review of studies that have attempted to summarise the factors associated with use comprehensively (Cousins & Leithwood, 1986; Johnson, 2009; Shulha and Cousins, 1997). This is followed by a discussion of the factors identified by other studies sorted by of they refer to evaluation implementation, stakeholder involvement, or program/organisational context factors. While some studies have attempted to rank the relative importance of these factors (e.g. Cousins & Leithwood, 1986), the review emphasises that each of these categories of factors are important to consider in evaluation planning.

### 3.3.1 Factors associated with the influence of evaluation

Some few studies have set out to understand the factors associated with evaluation influence, distinct from evaluation use. While these studies are much more significant to the present research as they employ the same conceptual framework, the emphasis in these studies is on evaluation for decision and policy making. This thesis has argued that human service practice is not analogous to decision-making, and that the factors associated with practitioner influence need to be explored in their own right. That said this section will summarise the factors in the literature associated with evaluation influence.

Oliver (2008) explored the factors associated with greater influence in 22 evaluations of disaster responses of International Non-Profit Organisations (INGOs). While exploring influence using the Henry and Mark (2003), the factors associated with influence were primarily drawn from the observations of participants. Looking
at influence overall, the researcher suggests timing was important, primarily the availability of the evaluation report for planning days and formal review times. Morale was also important, primarily in terms of providing a forum for INGO staff to reflect and discuss what went well and what could be done better. Interviews with people in a range of roles in the organisation identified some more specific issues that impacted evaluation influence: the length of reports, the distribution or availability of reports, being able to quickly identify the relevance for their work, and a clear sense of who was accountable to ensure the lessons of the evaluation are adopted. Practice level staff expressed a concern that evaluations could be too theoretical or academic, and felt that simple language was always better in reports. Participants also talked about the lack of a learning culture in terms of the attitudes and behaviour of individuals in the organisation.

While not specifically seeking to explore the factors associated with evaluation influence, the two related case studies carried out by Diaz-Puente, Yague and Afonso (2008), and Diaz-Puente, Montero, and Carmenado (2009) concluded that empowerment was a vital part of increasing evaluation influence, particularly in the context of communities with limited evaluation experience. What was initially seen as an imposition came to be seen as an empowering collaboration and a tool for continuous improvement. Similarly Morabito (2002), and Rebollos, Fernandez-Ramirez and Conton (2005) concluded from their case studies that evaluators should engage in a facilitative and collaborative approach in order to promote evaluation influence. From a study of evaluation influence in higher education, Fjellstrom (2007) emphasises the importance of deliberation and shared ownership, however cautioning that these can be threatened by a the lack of institutional support.
Drawing on Kirkhart’s (2000) definition of evaluation influence, Alexander (2003) examined three health service evaluations and explored the moderators of influence in each case. Three main barriers were mentioned by the participants: the qualities of the staff conducting the evaluation, the lack of funding for the implementation of recommendations, and the soundness and comprehensiveness of the evaluation. In addition to these barriers, the acceptance of the findings (i.e. if they confirmed pre-existing perceptions), and the ongoing relevance of the evaluation (i.e. if dramatic change had occurred between the evaluation and attempts to implement) were suggested as important moderators of the influence of the evaluation.

In a different approach Christie (2007) undertook a simulation study of evaluation influence, looking specifically at individual level influence, evaluation data types (large scale survey, case study, anecdotal evidence), and participant characteristics. Overall, participants were likely to be influenced by all data types in the simulated decisions they had to make, however those that were influenced by large scale surveys, were less likely to be influenced by anecdotes, and vice versa. Age, sex, and race did not have an effect on the influence of particular data types, however participants with a degree in education were less likely to be influenced by large-scale surveys. Where participants had existing beliefs about the effectiveness of a program, they were more likely to be influenced by survey data and less by anecdotal accounts.

3.3.2 Factors associated with the use of evaluation
Cousins and Leithwood (1986) and more recently Johnson et al. (2009) have undertaken meta-analyses that draw on the existing literature to put together a set of factors associated with the use of evaluation. Drawing on sixty-five empirical studies of use, Cousins and Leithwood (1986) clustered the factors associated with evaluation use into two categories, factors related to evaluation implementation, and
those related to the decision-making context. Table 3.1 displays these factors with the ranking of relative importance of each, which was arrived at by the prevalence of a relationship between the factor and use.

Table 3.1 Ranked Factors Associated with the Use of Evaluations (Adapted from Cousins & Leithwood, 1986, pp. 350-351)

<table>
<thead>
<tr>
<th>Rank</th>
<th>Factor</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Evaluation Quality</td>
<td>Methodological sophistication, type of approach to the evaluation problem, or the intensity of the evaluation activities</td>
</tr>
<tr>
<td>2</td>
<td>Decision Characteristics</td>
<td>Area of decision, decision context, significance of decision</td>
</tr>
<tr>
<td>3</td>
<td>Findings</td>
<td>Congruence with decision-makers’ expectations, value of the findings for decision-makers</td>
</tr>
<tr>
<td>4</td>
<td>Commitment to and/or Receptiveness to the Evaluation</td>
<td>Decision-makers commitment to the evaluation and attitude towards evaluation</td>
</tr>
<tr>
<td>5</td>
<td>Relevance</td>
<td>The extent to which the evaluation was geared towards the audience, if the evaluation was internal or external</td>
</tr>
<tr>
<td>6</td>
<td>Competing Information</td>
<td>Level of working knowledge, requirement for staff and service users input on decisions</td>
</tr>
<tr>
<td>7</td>
<td>Political Climate</td>
<td>Existing views, internal resistance to findings</td>
</tr>
<tr>
<td>8</td>
<td>Communication Quality</td>
<td>Communication style, ongoing communication activities, and breadth of dissemination</td>
</tr>
<tr>
<td>9</td>
<td>Information Needs</td>
<td>Intensity of information needs, type of information required, variance in audience needs for information</td>
</tr>
<tr>
<td>10</td>
<td>Personal Characteristics</td>
<td>Organisational role</td>
</tr>
<tr>
<td>11</td>
<td>Credibility</td>
<td>Function of the reputation, title, sex, evaluator’s belief in their work, decision-maker’s perception of the face validity of the evaluation, perceived appropriateness</td>
</tr>
<tr>
<td>12</td>
<td>Timeliness</td>
<td>On time for an important decision, or completed later than expected</td>
</tr>
</tbody>
</table>

As suggested in the introduction to this section, these studies focus on a view of evaluation use as an input to high level decision-making in an organisation, making it difficult to interpret the relative importance of these factors in practitioner learning. Yet, these provide an initial set of factors to explore, which are expanded on by subsequent research.

Reflecting the developments in the study of evaluation use highlighted by Shulha and Cousins (1997), Johnson et al. (2009) expanded on these to include a
number of new factors, primarily related to the way stakeholder involvement had been managed. Following Cousins and Leithwood (1986), there was a significant shift to the recognition of context in evaluation, and the importance of engaging the potential users of evaluation in order to promote use. Table 3.2 shows the additional factors identified by Johnson et al., (2009), which reflects the research focus on the role of stakeholder participation in fostering use. These additional categories reflect the direction of the evaluation research literature, which has been engaged in exploring effective stakeholder involvement.

*Table 3.2 Additional Factors Associated with the Use of Evaluations (Adapted from Johnson et al., 2009, pp. 383-385)*

<table>
<thead>
<tr>
<th>Category</th>
<th>Factor</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Evaluation Implementation</td>
<td>Evaluator Competence</td>
<td>Characteristics of the evaluator outside the evaluation process, cultural competence, leadership style</td>
</tr>
<tr>
<td></td>
<td>Commitment or Receptiveness to Evaluation</td>
<td>Stakeholder involvement creates a commitment or receptiveness to evaluation</td>
</tr>
<tr>
<td></td>
<td>Communication Quality</td>
<td>Stakeholder involvement promotes improved communication</td>
</tr>
<tr>
<td>Stakeholder Involvement</td>
<td>Direct Stakeholder Involvement</td>
<td>Direct link between involvement and use</td>
</tr>
<tr>
<td></td>
<td>Credibility</td>
<td>Stakeholder involvement increased the credibility of the evaluation process and/or evaluator</td>
</tr>
<tr>
<td></td>
<td>Findings</td>
<td>Stakeholder involvement in knowing about and understanding evaluation findings</td>
</tr>
<tr>
<td></td>
<td>Relevance</td>
<td>Stakeholder participation to build organisational context and concerns into the evaluation design</td>
</tr>
<tr>
<td></td>
<td>Personal Characteristics</td>
<td>Involvement of evaluation stakeholders at different levels in the organisation</td>
</tr>
<tr>
<td></td>
<td>Decision Characteristics</td>
<td>Involving a range of stakeholders in different settings (e.g. non-traditional bureaucracies) depending on the characteristics of the decision</td>
</tr>
<tr>
<td></td>
<td>Information Needs</td>
<td>Stakeholder involvement facilitated the introduction of their information needs</td>
</tr>
</tbody>
</table>

Exploring the relative importance of each of the factors identified by Cousin and Leithwood (1986), Burr (2009) undertook surveys and telephone interviews of
the grant project directors of undergraduate preparation programs. In contrast to the previous studies discussed, the most important factors for use in this research appear to be the relevance of the evaluation to the program, and project directors’ commitment and receptiveness to evaluation.

While illuminating, the factors identified in meta-analyses of studies exploring use are difficult to extrapolate to the context of practitioner learning. While some of the studies - particularly those of Johnson et al., (2009) - are concerned with use in the broader sense of learning or conceptual use (Cai, 1996; Lafleur, 1995; McCormick, 1997), the majority of the studies concern decision-makers and the organisational use of findings. It is an assumption of this research that practitioner learning represents a different process to organisational processes of decision-making (Weiss & Bucuvalas, 1980), and that different combinations of factors may be important for the difference change mechanism described in evaluation influence. The extensive evaluation use literature represents a significant and contemporary body of work to draw a list of possible factors from for exploration in this research.

Each of the recent studies reviewed examined a broad set of factors, these studies have been sorted into the types of factors their findings emphasise: evaluation implementation factors (how the evaluation was carried out); stakeholder involvement factors (how potential users of the evaluation are treated); and program/organisational context factors (the setting where the evaluation is to be used).

3.3.2.1 Evaluation implementation factors
Evaluators acknowledge the importance of evaluation implementation factors, along with some stakeholder factors in their practice. Drawing on the experience of 1,140 members of the American Evaluation Society, Fleischer and Christie (2009) investigated the most important factors that facilitated use according to evaluators.
The items with the highest level of agreement were planning for use at the beginning of the evaluation (91%), identifying and prioritising intended uses of the evaluation (89%), communicating findings to stakeholders as the evaluation progresses (87%), identifying and prioritising intended users of the evaluation (86%), involving stakeholders in the evaluation process (86%), and developing a communication and reporting plan (86%). This emphasis on implementation factors may be related to the lack of control evaluators have over the program and organisational context and their attention to the impact their practice can have on the likelihood of evaluation use.

Lawrenz et al. (2007) describes some of the advantages of different forms of dissemination for the use of evaluation, arguing that the approach needs to be suited to intended users. While not collecting any empirical data, the researcher suggests a set of factors that should be considered in developing dissemination strategies to maximise use amongst particular stakeholders (e.g. rationale, content, purpose, timing, development, distribution).

In an analysis of four case studies of program evaluations in private universities, Bober and Bartlett (2004) examined all of the factors described by Cousins and Leithwood (1986) and found evaluation implementation factors to be most important in determining evaluation use. Specifically communication quality, timeliness, quality, credibility, and relevance were found to be important across the different groups of stakeholders. This study emphasises the importance of the design of the evaluation in the likelihood of the use of the evaluation.

3.3.2.2 Stakeholder involvement factors
Karan (2009) explored the factors associated with the use of evaluations in the NGO sector through a survey of 111 staff from 40 NGOs in the United States, with his findings emphasising the importance of individual’s experiences and engagement with the evaluation. Table 3.3 presents the factors found to be important in
evaluation use, with Karan (2009) emphasising human factors as most important to fostering use. While the definition and understanding of *evaluation use* employed in this study emphasises evaluation for decision-making and organisational modes of learning, there is some relevance to practitioner learning.

*Table 3.3 Factors that Increase Evaluation Use (Adapted from Karan, 2009, pp. 153-160)*

<table>
<thead>
<tr>
<th>Category</th>
<th>Factor</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Human Factors</td>
<td>Intended Users</td>
<td>Involving intended users from all levels of the program decision-making hierarchy</td>
</tr>
<tr>
<td></td>
<td>Interests and Biases</td>
<td>Balancing different interests by planning and communicating intended uses</td>
</tr>
<tr>
<td></td>
<td>Professional Capabilities</td>
<td>Understanding administrative and organisational skills and building this into utilisation planning</td>
</tr>
<tr>
<td>Evaluation Factors</td>
<td>Evaluation Procedures</td>
<td>Stakeholder participation in the procedures of the evaluation</td>
</tr>
<tr>
<td></td>
<td>Substance of Information</td>
<td>The availability of information that is specific and relevant to users</td>
</tr>
<tr>
<td></td>
<td>Reporting</td>
<td>The effective reporting of findings that balance credibility with the messages for action</td>
</tr>
<tr>
<td>Organisational Factors</td>
<td>Organisational Culture</td>
<td>Inclusive and participatory decision-making, facilitated use, and organisational commitment to use</td>
</tr>
<tr>
<td></td>
<td>Routines and Processes</td>
<td>Systems and processes in the organisation that facilitate information sharing and interpretation</td>
</tr>
</tbody>
</table>

In his discussion Karan (2009) reflects on the importance of considering the experience of intended users of the evaluation in order to foster use, which includes the end users (human service practitioners). The factors identified reflect the agency of individuals and their capabilities in the use of evaluation, which also occurs within the broader context of organisational culture and the routines and norms in the organisation. This emphasis on the experience of the individual and its connection to organisational culture and processes factors suggest that stakeholder engagement is central to fostering *evaluation use*.

In a survey of 42 senior legislative officials in the United States, Vanlandingham (2011) found that evaluators who engaged regularly with
stakeholders and had actionable findings in their reports had more impact. In discussing the findings of the survey, he emphasises the importance of networks between evaluators and the potential users of the evaluation information, but highlights the tension for evaluators between providing objective findings and providing useful information for stakeholders.

Taut and Alkin (2003) undertook a study of the perceived barriers to the implementation of evaluation findings – presented as a specific type of evaluation use. In interviews with 18 staff members, Taut and Alkin (2003) categorised human factors as the major potential barrier to the implementation and use of evaluation by program staff. Specially, the lack of trust in evaluators by program staff and in the process in general, which the researcher attributed to inadequate relationship building. This finding is of particular relevance for this study as it focused on program staff members, who are akin to human service practitioners.

**3.3.2.3 Program and organisational context factors**

Lederman (2012) presents a model of how evaluation context and implementation may interact, suggesting that different combinations of the variables are conditions for particular types of change. While not specifically considering human service practitioners as end users – there was a focus on formal and documented decisions informed by the evaluation (p. 161) – this model has relevance to this research through the suggestion of necessary evaluation conditions (novelty value & evaluation quality) in particular organisational contexts (pressure for change & level of conflict). This attempt to simplify the complexity of evaluation use - a point Lederman (2012, p. 174) himself concedes – may in fact may be an effective way to develop knowledge about evaluation use in the decision-making context, but the focus on discrete decisions make this model ill-suited for considering the influence of evaluation among human service practitioners.
Balthasar (2009) emphasises the importance of how evaluation is institutionalised in organisations, while suggesting that stakeholder engagement is a pre-requisite for *evaluation use*. Following earlier work in developing a theoretical model for the effects of institutional design on *evaluation use* (Balthasar, 2006), the researcher went about examining how different institutional arrangements for evaluation impacted the use of 300 evaluations from the Swiss Federal Administration. Balthasar (2009) suggests that involving stakeholders and meeting their requirements is a prerequisite for the use of evaluation, but that different types of institutional arrangements seem to emphasise different types of use. He suggests that for general use, it is important that formative goals are pursued, or if summative goals are pursued they should be complimented by the embedding of evaluation across the organisation. Of particular relevance for this research, the Balthasar (2009) study suggests that external evaluations emphasise *conceptual use*, a form of use most akin to learning.

Murphy’s (2007) results emphasise the importance of a culture of organisational learning in order to increase the likelihood of *evaluation use*; she completed a survey of 283 non-profit organisations in order to examine the organisational factors that facilitate the use of evaluation. Drawing on established behavioural measures (Carman, 2005; Fine, Thayer, & Coghlan, 1998) a survey instrument was constructed to explore: internal learning, accountability, and image building (public relations). Most of the factors studied were significantly associated with use (Table 3.4); however, these factors impacted differently on the different types of use outlined. Of particular interest is the finding that organisational learning capability had a significantly stronger impact on the use of evaluation for internal learning, as opposed to accountability or image building. As Murphy (2007) points
out “nonprofits who have leadership committed to evaluation and an organizational culture that supports learning are more likely to use evaluation to benefit their organisation than simply for external purposes such as Accountability or Image Building” (pp. 224-225).

**Table 3.4 Factors Associated with Evaluation Use in NGOS (Murphy, 2007)**

<table>
<thead>
<tr>
<th>Category</th>
<th>Factors Explored</th>
<th>Factors Found Significant</th>
</tr>
</thead>
<tbody>
<tr>
<td>Organisational Characteristics</td>
<td>Size, age, budget, time, type of NGO</td>
<td>None</td>
</tr>
<tr>
<td>Organisational Learning</td>
<td>Managerial commitment, systems perspective, openness and experimentation, knowledge transfer and integration</td>
<td>Managerial commitment, systems perspective, openness and experimentation, knowledge transfer and integration</td>
</tr>
<tr>
<td>Stakeholder Engagement</td>
<td>Board, internal evaluator, internal staff, executive director</td>
<td>Board, internal evaluator, internal staff</td>
</tr>
<tr>
<td>Leadership Evaluation Characteristics</td>
<td>Experience with evaluation, evaluation philosophy and perception of evaluation practice</td>
<td>Leadership evaluation characteristics¹</td>
</tr>
</tbody>
</table>

¹This was explored as a single variable.

While a comprehensive and well-designed study, the use of a single person in the organisation as a survey respondent is potentially problematic. While certainly not a fault of the study, (multiple perspectives of 283 organisations would be extremely onerous), case studies have the advantage of multiple respondents to triangulate accounts, as well as allowing for extended observations of the impact of the evaluation rather than a single point in time.

Looking at how leaders moved evaluation beyond accountability and into organisational learning, Alaimo and Van Slyke (2006) concluded that program evaluation cannot be used for organisational learning without the support and commitment of leadership. From interviews with a stratified random sample of forty-two chief executive officers of human service organisations it was found that effective leaders integrated evaluation with mission, and used evaluation results in
strategic planning. This suggests that for evaluation to foster learning and program improvement, institutional support to foster a culture of evaluation is required.

Using a survey, Botcheva, White and Huffman (2002) found that the beliefs, norms and behaviours associated with a learning culture were correlated with systematic data collection and external funding. While working from a relatively small sample, they concluded that while there was a high level of awareness of the importance of evaluation, there were limited resources in these agencies for systematic data collection and capacity building. They also found that the resistance, fear, and negativity that can result from evaluation, seems to be symptomatic of the lack of a learning culture.

While the previous studies emphasise the importance of engagement, without adequate resourcing and leadership, change is unlikely to occur. Employing a retrospective case study approach with staff at two high school programs for gifted students, Avery and Van Tassel-Baska (2001) found that structural factors impeded the use of data by program staff. Program staff had a positive perception of evaluation and its role in questioning implicit practices and raising consciousness, and potential to act as a catalyst for program and organisational change. Participants however felt that the potential for evaluation to do these things was impeded by the lack of resources and lack of leadership support for program change. This finding emphasises an in-principle openness on the part of practitioners to use evaluation for learning and program improvement, but that appropriate support needs to exist.

3.4 Conditions for Practitioner Learning from Evaluation

The relevance of the above research to the specific context of practitioner learning presents a challenge, one that cannot be resolved through inferences. While this thesis has argued that practitioner learning is not analogous to decision-making, this
remains an assumption worth questioning. Recognising the difference in focus, the factors drawn from the literature above do not represent a model to test in the context of practitioner learning, but rather a set of possible factors that may come into play in the discussion of the way evaluation influence played out. Yet there is a need to synthesise these factors to compare to the factors found to be important to evaluation influence in the present research, and to provide some consistency in language.

These factors present as possible mediators or conditions for practitioner learning from evaluation, as examined through the evaluation influence framework (Mark & Henry, 2004). The sets of factors presented in Figures 3.1-3.3 represent a distillation of the possible factors drawn from the existing literature that may mediate different evaluation influence mechanisms. While this research will not draw on these factors directly in the analysis of the conditions and barriers to practitioner learning, reviewing the existing literature serves as an important background. The conditions and barriers for practitioner learning will be explored, and then compared to the factors associated with evaluation use (Figures 3.1-3.3).

Drawing on the factors identified above, Figure 3.1 represents a map of the evaluation implementation factors that could potentially influence practitioner learning from evaluation. While many of the additions to Cousin and Leithwood's (1986) original categories appear to merely restate or reword these factors, each provides a greater level of conceptual depth. By focusing on practitioner learning from evaluation, these factors can be given a greater level of context, for instance credibility may be interpreted differently by practitioners from a variety of discipline backgrounds as opposed to decision-makers and policy staff.
Many of the evaluation implementation factors are likely to be determined by individuals' experience of the evaluation. Figure 3.2 outlines the factors related to stakeholder involvement. Recognising the importance of how stakeholder involvement is managed, Johnson et al. (2009) suggests this as a separate category of factors interrelated to evaluation implementation. These factors, while similar to those in Figure 3.1, emphasise the experience and perceptions of stakeholders.
While somewhat out of the control of an evaluator, institutional norms and the organisational context play an important role in *evaluation use*. While the studies identifying these factors have focused on the decision-making context, they represent a rough map of factors to explore in the practice context.
3.5 Chapter Summary

This chapter provides a definition of the practice learning the research has set out to explore, a description of the different roles evaluation can serve in learning and their connection to individual practitioner learning, and summarises the factors in the
literature associated with *evaluation use* in order to consider how they might apply to practitioner learning. This thesis takes the view that while organisational learning systems can set the values around learning, and even disseminate legitimised information, ultimately learning is an individual process of “self-reflection and questioning” (Antonacopoulou, 2001, p. 328). Practitioners reconcile evaluation information with their experiential knowledge and engage in a thoughtful process of reflection and questioning. The literature suggests that evaluation can play a number of roles in organisations that can connect to individual learning: providing outsider advice (Buckmaster, 1999; Ebrahim, 2005a; Owen & Lambert, 1995; Schwandt, 2005) the examination of implicit assumptions (Owen & Lambert, 1995; Preskill, 1994; Shaw & Shaw, 1997), the promotion of evaluation culture (Schwandt, 2005; Shaw & Shaw, 1997; Torres & Preskill, 2001), and the empowerment marginalised groups (Ebrahim, 2005b; Edwards, 2002; Hoole & Patterson, 2008). Much research has been done on the factors associated with *evaluation use* in the context of decision-making; however it is unclear how these factors might relate to practitioner learning. These include factors related to the implementation of an evaluation (Figure 3.1; quality, relevance, credibility, findings, timeliness, communication, competence), how stakeholder involvement in the evaluation is managed (Figure 3.2; communication quality, direct stakeholder involvement, credibility, findings, relevance, personal characteristics, commitment to or receptiveness to evaluation, decision characteristics, information needs), and the program or organisational setting (Figure 3.3; competing information, personal characteristics, commitment or receptiveness to evaluation, program characteristics, information needs, political climate, organisational learning, leadership).
4. CHAPTER 4: METHODOLOGY

4.1 Introduction
The aim of this research is to add to the body of knowledge about how evaluation can best influence practitioner learning, with a view to both program level change and learning that leads to broader social improvement. In other words, what are the conditions and processes that help or hinder learning and what are the implications of these for efforts to improve the influence of evaluation? These aims were central in shaping the design and approach of the research. This chapter will attempt to lay out some of the methodological choices I have made and offer some reasoned discussion of these choices, while positioning the research within a clear philosophical paradigm that acknowledges the capabilities and limitations of knowledge to represent reality (Guba, 1990). In the interests of demonstrating validity and methodological integrity, I will strive for a detailed and replicable description of the methods and procedures of the research. The chapter will begin with a discussion of the paradigms of inquiry implicit in the research approach, then go on to describe the methodology employed, and the procedures undertaken.

4.2 Criteria for Methodological Decision-Making
Cresswell (2005, pp. 21-23) and Patton (2002, pp. 230-234) offer similar lists of considerations that should factor into decisions about research approach. These include: first, the audience of the research, secondly the researcher’s view of the world influenced by their training, academic discipline, and personal experiences; thirdly, practical considerations related to the research context, and finally the fit between the research problem and the approach. As a research thesis, the potential
audience is a less important determinant of approach than research that is commissioned. As is often the case with a research thesis, the intended audience is limited to supervisors, reviewers, colleagues, and perhaps other research students interested in some aspect of the study.

A qualitative approach allows the participants to speak for themselves and to share their perspective, which depending on the manner of the acquisition and use of the data, there is a potential for qualitative research to be empowering for participants (Fetterman, 2002). In previous research there seems to be a preference for qualitative research amongst practitioners in the NGO sector (Cortis, 2006), particularly as there is a sense that such workers feel marginalised within their organisations and in the increasingly complex interrelationships between governments and NGOs (Carman & Fredricks, 2008; Ebrahim, 2005a; Meagher & Healy, 2003; Taut & Alkin, 2003). Particularly considering the hostility within the sector for top-down management oriented performance measures, and the view that these measures dehumanise and minimise the work of practitioners (Healy, Meagher, & Cullin, 2009; Meagher & Healy, 2003; Mintzberg, 1996; Stepney, 2000), practical considerations seem to suggest a qualitative approach.

The primary basis for the methodological choices of this research is the match between the problem and the approach, which also link into key ontological, epistemological, and methodological choices. Reflecting on the previous review of the literature and summary of the research problem, the proposed research has the following characteristics that are formative to the research approach I am proposing:

- The research is concerned with the meaning of a particular phenomenon as constructed by the participants (in order to understand how evaluation influences their practice);
The data required to address the aims of the research will be rich and subjective accounts that will likely be inseparable from context and from the participant’s own experiences and personal biography;

- The research concerns the shared patterns of behaviour and implicit norms associated with culture groups (i.e. organisations and practitioners);

- The research will not seek absolute or objective answers to the research questions, only the summary of subjective responses and the patterns and relationships of meaning (Moustakas, 1994) amongst these responses;

- The researcher will be actively involved in the interpretation and construction of meaning in the data; and

- Theory fits into the research as a lens to direct the inquiry towards the themes and issues prominent in the literature.

These characteristics describe the nature of the present research problem and in turn prescribe some limitations to the suitability of particular research approaches. In particular, the focus on the experiences of participants and the lack of an absolute or objective answer to research questions suggest that the most appropriate methodology is likely to be a qualitative approach.

### 4.3 The Qualities of Qualitative Research

Qualitative research involves examining the nature of phenomena from the perspective of the people involved, who bring their own values and meanings to the events (Denzin & Lincoln, 2003). Qualitative procedures are interpretative, interactive, humanistic, and view social phenomena as complex and holistic (Rossman & Rallis, 1998). The key qualities of this approach are that qualitative research is: *participant oriented*; the process or *meanings of actions* are as important as the outcomes; the inquiry process is *open and adaptable* to discovery; and the *researcher is the instrument*. 
Research questions concerned with human perceptions and experiences require a research methodology that is able to represent the dense details of daily life (Schwandt, 2003, p. 294) and attempt to make sense of phenomena through the meaning the participants bring (Denzin & Lincoln, 1994). Qualitative research involves immersion in the everyday life of the participants, interacting with their perspectives and meaning (Marshall & Rossman, 1989), capturing the sense and significance of phenomena by studying actions and experiences in their natural state. The focus is on the perceptions and experiences of participants and the sense they make of their lives (Fraenkel & Wallen, 1990; Merriam, 1988) in order to understand the multiple and often conflicting perspectives of reality that exist (Lincoln & Guba, 1985). The onus is on the researcher to represent and interpret the experiences of participants with fidelity, and to foster an ethical and mutual exchange as the researcher attempts to reconstruct the participants’ realities (Lincoln & Guba, 1985; Merriam, 1988).

Qualitative research is well suited to exploring the meanings of actions and outcomes, but also the process or context of these phenomena, and how and why they occurred (Fraenkel & Wallen, 1990; Merriam, 1988). Importantly, qualitative research has the capacity to explore the socio-cultural conditions and motivations of actions, without which the meaning or value of particular practices or actions may remain elusive (Habermas, 1988). Qualitative research also has the advantage of being able to capture the lifecycle of a phenomenon (Schofield, 1990), rather than the kinds of snapshots collected by quantitative research. These qualities make qualitative approaches well suited to research questions that aim to make some useful attribution from the experiences or perceptions of participants (Patton, 2002).
Qualitative research is naturalistic in that the researcher does not manipulate the setting; the research is discovery oriented (Guba, 1978) and places a premium on not constraining the outcomes or categories of the analysis (Patton, 1990). It is an investigative process where researchers gradually make sense of a social phenomenon by contrasting, comparing, cataloguing and classifying the object of study (Miles & Huberman, 1984), and observing the variations and similarities of the reported experiences (Rossman & Rallis, 1998).

**4.3.1 The validity and reliability of qualitative research**

Quantitative research, which operates within the logical-positivist paradigm, is concerned with the testing of hypothetical generalisations and causal relationships (Cresswell, 2005; Hoepfl, 1997). The criteria by which quantitative research is judged are validity (internal and external) and reliability, characteristics somewhat at odds with qualitative research and social constructivism (Lincoln & Guba, 1985). These concepts have in turn been reinterpreted (Strauss & Corbin, 1990) in order to better reflect the characteristics of ‘good’ qualitative research. In arguing for the legitimacy of qualitative approaches and naturalistic inquiry as research, Guba and Lincoln (1989) propose alternate criteria to judge the quality of qualitative research based around trustworthiness as a surrogate for validity and reliability. Trustworthiness as criteria for quality is described as the case that the findings are “worth paying attention to” (Lincoln & Guba, 1985, p. 290), or that the research is a faithful reconstruction of the definitions or meanings present in the research context (Guba & Lincoln, 2005, p. 257). In place of the criteria associated with quantitative research, Guba and Lincoln (1989) offer:

- Credibility (in place of internal validity): Whether the findings represent a credible conceptual interpretation of the raw data (e.g. interviews and observations);
- Dependability (in place of reliability): An assessment of the quality of the processes of data collection, data analysis and theory generation, and the stability of data over time;
- Confirmability (in place of objectivity): How well the findings are supported by the data collected and not influenced by the perspective of the researcher; and
- Transferability (in place of external validity): The degree to which the findings can apply or transfer outside the bounds of the research (p. 296).

Despite some disagreements in terminology (Lincoln & Guba, 2000) and some debate about the relevance of generalisability in qualitative research (Stenbacka, 2001), these criteria have remained central in discussion of the quality of qualitative research. I will refer back to how these criteria have factored into my methodological decision-making in the course of presenting the logic and method of the present research.

Qualitative research involves the researcher as the primary instrument in data collection, so validity depends on the skill, competence and meticulousness of the researcher (Patton, 1990). While there is inevitably a loss in rigour from the fallibility of the human instrument, Guba and Lincoln (1981) argue that flexibility, insight, and the ability to build on tacit knowledge offset this loss. This active role makes the researcher relevant to the knowledge acquired (Banister, Burman, Parker, Taylor, & Tindall, 1994); with an awareness of the gap between the research object and the way the object is represented through the interpretations of the researcher. Because of the involvement of the researcher in both the collection and interpretation of the data, qualitative research should demonstrate what Patton (1990) calls “empathic neutrality” (p. 41). Rather than demanding objectivity, empathic neutrality requires that the researcher be concerned with understanding the world in its
complexity, be aware of the content of his/her own interpretations, and to take a neutral non-judgemental stance to the data (Patton, 1990).

4.4 Paradigms of Inquiry

In elucidating a framework for understanding the legitimacy of knowledge claims made through research, Guba and Lincoln (1994) suggest that three questions need to be addressed:

1. The ontological question: What is the form and nature of reality and, therefore what is there that can be known about it?
2. The epistemological question: What is the nature of the relationship between the knower or would-be knower and what can be known?
3. The methodological question: How can the inquirer (would-be knower) go about finding out whatever he or she believes can be known? (p. 108)

Each paradigm that purports to make legitimate claims to knowledge needs to offer some answer or philosophical position to each of these questions. The well-trodden path of particular research approaches or paradigms bring with them implicit or taken for granted answers to Guba and Lincoln’s (1994) questions. However, adopting a pragmatic approach requires some effort in considering the assumptions of the research, as they are determined more by what is appropriate for the research problem (Patton, 1990), than a firmly held position one is particularly invested in.

4.4.1 Ontology

Ontology is the philosophy concerned with what exists in the context of the social, or what is the nature of social reality. Different paradigms make claims about the world underpinned by these assumptions, and claims about what exists, the conditions of their existence, and the ways they are related (Crotty, 1998). These suppositions then form the basis of how a researcher approaches the research problem. Ontology tends to be discussed relative to a continuum, with realism on one end and idealism on the
other (Blaikie, 2007). Because the positions on this continuum tend to be tied to rigid systems of philosophy, there is a preference to avoid questions about reality and the laws of nature at the core of pragmatism (Cresswell, 2005) and to focus on what works in the context of the inquiry. However, it is not sufficient to disregard ontology in favour of epistemology as the two are inexorably linked. To this end I will attempt to lay out an ontological position for this research influenced by depth realism and its lighter counterpart, pragmatic realism.

Realism is associated with the positivist or scientific worldview, and suggests that there is a single reality governed by physical laws (Orlikowski & Baroudi, 1991). However depth realism takes the position that reality consists of three levels or separate ontologies: the empirical, the actual, and the real (Bhaskar, 1978). The empirical is the world experienced through our senses, the actual includes events whether or not anyone is there to observe them, and the real is the underlying processes that generate events. Like the shallow or conceptual realist (Blaikie, 2007), this position assumes that a single reality exists, however this is manifested in unobservable structures and processes whose influence can only be inferred. While traditionally, knowledge is derived from the recording and observation of the empirical, Bhaskar (1998) emphasises the importance of investigating and understanding the causes of things, the unseen mechanisms or processes that produce the empirical observation. Houston (2001) provides the example of the arrangement of metal filings when subjected to magnetic forces; the empirical observation is the patterns and movements of the metal filings, however the unseen mechanism at work and the only satisfactory explanation for the phenomena is magnetism.

Unlike Houston’s (2001) example from the natural sciences, the application of depth realism in social science research relies on human participants and a human
interpretation of these experiences. Observations of the empirical are unavoidably subjective and formed by the individual’s interaction with language, culture, and experience (Phillips, 1987). In the present research, which takes place within an organisational context, there are additional institutionalised norms and values that are likely to also influence the perspectives of participants (Lippi, 2000). Depth realism provides only a transitive view of the world that is embodied with the values, prejudices, and ideology of the socio-historical context; a fully accurate picture of the social world is not possible or even desirable (Bhaskar, 1991).

Depth realism further diverts from other forms of realism by suggesting that the world represents an open and irreducible system, filled with more mechanisms than one could ever hope to measure or control (Bhaskar, 1998). For the inquirer, their role is to investigate some of the mechanisms at play in order to be able to say something about the outcomes that tend to occur under certain conditions, and to develop thinking around explaining and understanding these tendencies (Houston, 2001).

Pragmatic realism represents a broad spectrum of thought about truth or reality, bound only by a focus on defining reality as a means-to-an-end to the task of producing some useful insight. From this perspective, ontology is referred to as a grammar that allows for the understanding and description of phenomena (Harre, 1997). Pragmatists are concerned with addressing human problems and attempt to avoid the abstract reasoning employed in other ontologies. Kiven and Piironen (2004) describe both realists and idealists as playing a kind of esoteric language game that is divorced from the problems and concerns of human life. By emphasising ontological approaches that are fit for the purpose of understanding and
resolving these problems, pragmatists attempt to remain agnostic to any particular fixed position on ontology.

The present research reflects a depth realist (Bhaskar, 1978) perspective of ontology, however with a degree of scepticism for abstract notions of reality influenced by pragmatic realists such as Rorty (1990). I assume that the empirical data collected in the course of the research has a link to reality, constituted by the causal mechanisms underlying the experiences of the participants. Along with this assumption is the understanding that the mechanisms that operate in the social world are more numerous and complex than a researcher could ever hope to describe, therefore the research is limited to the identification of tendencies rather than any certainties in the data. In addition, the nature of relying on the human reporting of experiences in the social sciences means that the link between the empirical and the real is even more distant, as the subjective perspectives of the participants add an additional layer of complexity. For the purposes of the research problem this ontology allows for the study of the mechanisms at play in attempts to foster learning from evaluation, fulfilling the aims of the research by offering a means by which to improve understanding of how to maximise the influence of evaluation in creating learning.

4.4.2 Epistemology
Epistemology concerns how knowledge can be said to be accurate or legitimate (Crotty, 1998), primarily concerning the relationship between the researcher and the object of interest (Guba & Lincoln, 1998). The primary epistemological influence of this research is social constructionism, which emphasises that knowledge is constructed through social actions and discourses embedded in the social context (Schwandt, 2007). Crotty (1998) identifies three key assumptions of social constructionism:
• Meanings are constructed by people as they engage with the world;
• People engage with and make sense of the world from a particular historical and social point of reference; and
• Meaning arises out of interaction with a human community.

Social constructionism is particularly relevant in studying experiences within organisations, as within institutions social knowledge is actively constructed “against a backdrop of shared understandings, practices, language and so forth” (Schwandt, 2003, p. 197). Kwan and Tsang (2001) and Mir and Watson (2000) discuss the development of constructivist research approaches in the study of organisations, observing that the need to understand organisational complexity has driven a break from the traditionally positivist approach in the discipline.

While social constructionism is commonly associated with a radical relativist position (Kwan & Tsang, 2001); constructionism contains a spectrum of epistemological orientations. I have come to accept as Kwan & Tsang (2001) and Schwandt (1994) suggest, that there is no internal inconsistency between realism and the study of contextualised social constructions, merely that the constructionist worldview requires the co-construction of meaning, the reasoning for how the co-creation enhances the research process, and critical subjectivity on the part of the researcher.

A key aspect of social constructionism is that actors in the social environment are involved in the construction of reality through their own interpretations of experiences and the meaning they ascribe to the actions of others (Cresswell, 2005). Constructionism assumes that it is impossible for the human senses to represent this world without the encumbrances of the human context and condition. Merely by observing or exploring these constructions, the researcher is entangled and inseparable from them. Co-construction fulfils not only an ethical
prerogative to engage with participants, but also a way of effectively engaging with bias and perspective on the part of the researcher (Brydon-Miller, Greenwood, & Maguire, 2003). This process is described by Charreire-Petit and Huault (2007) as a “...dialectic and iterative process built around analysis, criticism, reiteration, and reanalysis” (p. 78), presupposed by trusting and conversational interactions between the researcher and participant.

This research is concerned with studying the experiences and perceptions of individuals, and the meaning they construct from these experiences. Engaging in co-construction with participants enhances the research process by allowing the participants to reflect on their own responses and to provide feedback on the interpretations of the researcher. Particularly as the intent of the research is to identify patterns and relationships of meaning in order to make some observations about possible underlying mechanisms and processes, co-creation represents an important grounding for the research.

Through the interaction of the investigator and the object of interest, the researcher brings their own standpoint to the task of constructing knowledge from the experiences of participants (Denzin & Lincoln, 2000) in the context of shared interpretations, practices, and language, and in socio-cultural ways of being (Blaikie, 2007). This task requires a degree of reflexivity, which Pollner (1991) defines as “an ‘unsettling,’ i.e., an insecurity regarding the basic assumptions, discourse and practices used in describing reality” (p. 370). In addition to questioning the truth claims of others, researchers need to recognise the philosophical basis and internal logic implicit in the way they construct meaning (Alvesson, 2003). The social constructionist perspective on reflexivity in this process is one of “betweeness” (Cunliffe, 2003, p. 986), where social realities are constructed between the researcher
and participant in conversation (Shotter, 1993; Watson, 1994). In accepting the intersubjective nature of their research, the investigator needs to explore how meaning is constructed between themselves and participants, and to recognise how we as researchers construct meaning through our own taken for granted suppositions, actions, and linguistic practices (Cunliffe, 2003).

4.5 Methodology

The research concerns two key research questions, which the methodology has been developed to address:

- What are the conditions and barriers to human service practitioner learning from evaluation in human service programs delivered by NGOs; and
- How could evaluation be undertaken in order to maximise practitioner learning?

Considering the research problem and the ontological and epistemological assumptions mentioned previously, the discussion turns to the methodological approach to collecting and analysing data, and the nature of the knowledge that will result from this process. I will begin by discussing the research strategy employed and the style, internal logic, and theoretical grounding that the approach brings with it. This will further lead to a discussion of the specific methods that were employed and the role of each in the research strategy. The role of the researcher and a description of the procedures undertaken in the course of the inquiry will follow.

4.5.1 The abductive research strategy

Abduction is a mode of reasoning combining both deduction and induction in order to make logical inferences about the world. Given (2008) presents this mode of reasoning in the form: *Some event, X, is surprising to us. But if some explanation, Y, were in place, then X would be ordinary. Therefore, it is plausible that X is actually a case of Y* (p. 1).
Blaike (2003) describes this as a process of attempting to discover the underlying mechanisms that satisfactorily explain observed regularities. In the interpretive social sciences, abductive logic is employed towards the generation of theory from social actors’ language, discourse and practices (Blaikie, 2007). This process begins with the first order task of describing the meanings and interpretations of the tactic, symbolic and mutual knowledge of the participants, then the application of social scientific concepts and theory to this knowledge (Bhaskar, 1979, 1986). Blaike (2007) summarises the layers or depth of meaning from this process as follows:

*Everyday concepts and meanings provide the basis for social action/interaction about which social actors can give accounts from which social scientific description can be made from which social theories can be generated or which can be understood in terms of existing social theories or perspectives* (p. 90).

This research will utilise the abductive research strategy to collect accounts of the participants’ experiences and perceptions of evaluation and the implicit understandings that form the basis of their actions. This forms the basis for the generation of social theory and the application of existing theories and perspectives about the social setting (human service organisations) and the social phenomena (practitioner learning from evaluation). The aim of this research is for these theories and perspectives to add to the knowledge of practice level learning through evaluation in non-profit organisations and to produce some practical insights into how to improve evaluation practices in a way that leads to social improvement.

This research was undertaken as two parts, although the data was collected concurrently: (a) case studies of how the evaluation played out, (b) the participant’s experience of the evaluation that provide an explanation of why the evaluation was
influential, and how participants thought it could have been more influential. The

case studies serve an important role in describing the narrative of the evaluations and
how it fits into the wider story of the program and organisation. The case studies are
instrumental in the sense that they are undertaken with the aim of assisting the
understanding of the phenomenological accounts provided from the interviews
(Stake, 2003). The interviews that serve as the primary data of this research are
phenomenological in the sense that they concern “...the structure and essence of
experience of phenomenon” (Patton, 1990, p. 69). These detailed accounts will then
be the source material for the observations and theory that explain the observed
regularities.

4.5.2 Part 1: Case study approach

The aim of the case studies in the present research is to provide detailed description,
and practical insight into the experiences of evaluation obtained through the
interviews. This form of case study, which Stake (2003) describes as a simple case
study, provides data that is descriptive and serves to present a variety of perspectives
as well as a means of triangulating these perspectives to some likely course of events.
Case studies represent a means of generating and testing theory within a single
bounded system (Smith, 1978; Yin, 2009), and are particularly useful in examining
rich phenomena where the boundaries between the phenomena and the context are
not clear (Yin, 2009).

The case studies serve as the narrative of the evaluation; including the context
that existed prior to the evaluation, the impetus for the evaluation taking place, the
events that constitute the evaluation, to everything that occurred following the
evaluation. The case study is presented in narrative form, emphasising the record of
events. A case study protocol (Appendix A) was developed to help set out the scope
and manner of the inquiry, allowing some consistency in the types of data and the
procedures undertaken within each organisation. However, there was considerable variation from the planned procedures of the research; the implications of these variations for the credibility, dependability, and transferability of the collected data will be discussed in the context of the case study.

**4.5.3 Part 2: Phenomenological approach**

The interviews and the analysis of the interviews are oriented towards understanding the subjective or lived experience of participants concerning the phenomenon of learning resulting from evaluation. While hardly a radical notion in qualitative research, phenomenology brings with it a particular philosophical style (Giorgi, 1997) or attitude toward the world. This attitude is characterised by the goal of attempting to obtain a deep understanding from the participants’ everyday experiences (Broussard, 2006). This deep understanding from the lived experiences of a group of people has been characterised by Banaga (2000) as the attempt to find the underlying structures within experiences.

The phenomenological method has two main approaches with differing philosophical bases. It should be clear by a reading of the epistemology and ontology sections that an interpretative or hermeneutic approach will be adopted (Cohen, 1987), rather than a descriptive or eidetic phenomenology. Hermeneutic phenomenology developed from and directly challenged some of the assumptions of the descriptive approach, primarily in going beyond description to look for meaning embedded in the understanding and everyday experiences of people (Lopez & Willis, 2004; Racher & Robinson, 2003). Some of the main features of hermeneutic phenomenology include:

- A focus on human experiences that participants are not always conscious of (Solomon, 1987), which is to say “... meanings that are
not always apparent to the participants but can be gleaned from the narratives produced by them” (Lopez & Willis, 2004, p. 728).

- The experiences and perceptions of individuals captured in their narratives contain knowledge about what that person experiences in everyday life. The relationship between individuals’ realities are influenced by the lifeworld they occupy (Heidegger, 1962).

- The concept of situated freedom (Leonard, 1999), meaning that while individuals are free to make choices, these are linked to and circumscribed by social, cultural, historical, and political contexts (Heidegger, 1962; Leonard, 1999; Lopez & Willis, 2004). A hermeneutic phenomenological inquiry focuses on the “meanings of the individuals’ being-in-the-world and how these meanings influence the choices they make” (Lopez & Willis, 2004, p. 729).

- As opposed to descriptive phenomenology, a hermeneutic approach assumes that pre-suppositions, knowledge of the research literature or pre-existing theory are valuable in the inquiry and cannot be eliminated (Lopez & Willis, 2004). Researchers should make their preconceptions, assumptions, and their frame of reference explicit and provide some evidence that these did not bias the interpretation of the narratives (Lopez & Willis, 2004).

- Inter-subjectivity or co-constitutionality (Koch, 1995) are part of the interpretative or hermeneutic research process; in that the meanings arrived at through the research are a blending of the meanings of both participant and researcher. Research of this kind is judged by the logic and plausibility of the research findings and by the compatibility of the findings with the realities of the participants (Annells, 1996).

Utilising a hermeneutic/interpretive phenomenological approach, this part of the research will aim to draw on participants' accounts of how the evaluation influenced their practice, and the factors that determined or mediated the influence of the evaluation. This approach draws on Mark and Henry’s (2004) evaluation influence framework, in order to understand how the evaluation has influenced
practice and the program, followed by the participants’ experiences that suggest why
the evaluation has been influential. By looking at specific influence events, as well as
participants’ overall experience and perception of evaluation, this research aims to
generate theory concerned with how to undertake influential practice oriented
evaluation (Schwandt, 2005). The research will also question the assumption that the
kind of practice learning this research is concerned with is substantially different
from other forms of use/influence studies in the evaluation research literature. The
research will reflect an awareness of the inter-subjective nature of this type of
inquiry, with the researcher employing reflexivity to manage the biases that are
inevitably part of employing the human instrument in research.

4.6 Organisation/Program Selection

The research involved two NGOs that provide social services within the greater
Sydney metropolitan area that had been evaluated in the past twelve months or had
an ongoing evaluation expected to be completed in the twelve months following the
commencement of the research. Setting out with the idea to examine practice level
learning from evaluation, and being new to Sydney, I started by meeting with
research and evaluation staff in the major NGOs operating in New South Wales. In
these unrecorded and off the record interviews, I was able to find out how the
challenges of implementing evaluation played out in the socio-political context of
human service NGOs. Along with developing my knowledge of the context, the
interviews served to develop rapport with the gatekeepers of research in these
organisations.

Another aim of these early interviews was to identify suitable programs to
include in my research. I set out with a number of considerations for the type of
programs I was looking for to increase my chances to observe the phenomena of
interest; practitioner learning from evaluation. First, I focused on human service programs, programs with staff that identify with a particular set of goals or ideals within the context of the super-ordinate goals of the organisation (Wholey, Hatry, & Newcomer, 2004). Secondly, on-going sustained programs with the capacity to attempt to improve practice, such as those with reasonably secure funding and support for their continuation. Thirdly, programs that receive most of their funding from a state or federal government agency; that is the NGO is delivering a program that has been funded or 'purchased' (Darcy, et al., 2009; O'Shea, 2007) and is responsible to that agency for their outputs and outcomes. Fourthly, programs that involve social services constituted by a human interaction between staff and clients in a community setting. Lastly, that the evaluation had an explicit goal of fostering internal learning and improvement within the program and organisation. The selection of programs with these characteristics increased the chances of encountering examples of learning from evaluation, allowing the research to explore the factors that are critical to this kind of influence taking place.

The selection of these sites were based on Patton’s (1990, pp. 182-183) strategies of purposeful sampling. Within the bounds of the criteria listed above, the intent was to include cases that were typical (Patton, 1990, p. 171), which is to say typical of the types of programs that undertake evaluation oriented towards learning. While it might be problematic to extrapolate the findings to other types of programs and organisation, the characteristics of the organisation are part of the scope of the inquiry, indeed practice/organisation context represent half the factors associated with evaluation use/influence in the literature. By seeking to explore practitioner learning from evaluation in what seems to be an ideal environment (i.e. evaluation with an internal learning agenda, organisation with capacity to participating in and
facilitate practice learning), the aim is to develop theory that can be tested and applied more generally.

The initial interviews were formative in terms of my understanding of the NGO context. I had an initial discussion with one organisation that described the challenges of operating with a large volunteer workforce, meaning that the organisation was run very informally and there were many groups and sectors with limited contact with the head office. The implication of this was that despite having a developing research and evaluation capacity, there was often difficulty in obtaining data from the different arms of the organisations. Despite opportunities for increases to funding and services, the level of data collection and evaluation required was simply not possible at this organisation due to the political climate created by the reliance on volunteers.

The second organisation I spoke was The Benevolent Society. The organisation had significant evaluation capacity, which was used primarily to conduct internal evaluations, along with supporting the implementation of externally conducted evaluations. Almost all of their services were directly funded by state and commonwealth agencies, with almost all of them having some requirement for evaluation. However, most of these are focused on output data for accountability purposes. They also engaged in conducting internal evaluations that emphasised a formative approach in order to foster internal improvement. These evaluations tend to be initiated internally as a response to some problem or issue, or ahead of attempts to acquire additional funding or support for particular programs.

From this initial discussion a number of suitable programs that had been evaluated recently were suggested. I spent some time looking into each of them, but

---

4 The organisation has a preference to be referred to as The Benevolent Society (with a capital T).
had a strong sense that one in particular met the sample inclusion requirements. I returned a few weeks later to have a more detailed discussion about this program with the head of evaluation who was responsible for the internal management and implementation of this external evaluation. With the program selected the next step was to obtain permission from the organisation to proceed with the research. This involved the preparation of a research approval form and presenting the proposed research to the organisation's research approval council. I presented for about fifteen minutes and then responded to follow up questions. Permission to undertake the research was granted, and contact was made with the manager of research and evaluation to begin recruiting the key respondents needed.

Similar to the second, the third NGO (Uniting-Care Burnside) I selected for inclusion had considerable evaluation capacity, with a mix of internal and external evaluations ongoing in a large and complicated organisational structure. While many of the programs delivered by this organisation were funded by Community Services NSW and the Commonwealth Department of Families, Housing, Community Services and Indigenous Affairs (FaCSIA), the head of research and policy emphasised that there was a mix of many different types of programs and funding arrangements. A number of programs were suggested, which I worked on researching following our initial meeting. Following the meeting I took a particular interest in a family support program that was undergoing a restructure as a result of an evaluation. Despite practitioner and community commitment and goodwill toward the program, there was a sense that the program overlapped a number of existing services and had limited effectiveness. Expecting that the evaluation would be implemented to adapt the program to fit the community need, which would involve considerable program and practice level changes, I selected this program. On
following up I was told that the program was being shut down that day, despite the availability of funding for the program; which could well be attributed to the evaluation. Taking note of this and the possible effects it may have on practitioners' perceptions of evaluation, I selected another program of interest. The process of getting permission to undertake the research was much more informal, I submitted copies of my university ethics clearance, and a rewritten version of the summarised proposal I submitted to the previous organisation. Following this I was allowed to proceed with my research.

I also made contact with a fourth organisation with considerable research and evaluation capacity. Following an initial discussion which yielded a number of programs of interest, there was some difficulty in following up to the point of getting permission due to concerns that program staff would be too busy around the time I proposed to conduct the research.

It should also be noted that the two case studies are not equivalent in terms of the timeline of study, and the number of participants. While the reasons for this will be described in more detail in the case studies, it is important at this point to note that these cases are not equivalent. The Benevolent Society case study, while protracted, followed the case study protocol closely. The Burnside case study became increasingly difficult to follow up as time went by, primarily due to organisational delays in releasing the final evaluation report, and as the organisation moved into a particularly busy time with the NSW social bonds trial (The Centre for Social Impact, 2011). While I attempted to postpone my interviews in order to be able to capture the impact of the dissemination efforts, I ran out of time and accepted that I would not be able to capture the same timeline of events as in the Brighter Futures evaluation. The decision was made to work with a final data collection timeline
ending in October 2012. As such, the Brighter Futures evaluation can be thought of as the main study and source of influence events, with the St Marys case offering complementary insights about how evaluation practice affects the potential for practitioner learning from evaluation.

*Figure 4.1* Timeline of the Evaluation in Relation to the Case Study Period

<table>
<thead>
<tr>
<th></th>
<th>2006</th>
<th>2007</th>
<th>2008</th>
<th>2009</th>
<th>2010</th>
<th>2011</th>
<th>2012</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brighter Futures</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>St Marys</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The grey gridlines indicate the duration of the evaluation, the duration of the case study in shown in black.

Over the course of research, two sector-wide developments occurred that changed the context of these programs. The first, which has already been alluded to is the trial of social bonds; a scheme where private equity groups can fund social programs and receive a benefit based on future cost savings to governments. In principle the idea is an appealing one, Hems and Flatau (2011b) point out that social bonds represent a middle ground between corporate philanthropy and ethical investing. The challenge is in measuring social impact in a meaningful way, and in managing any possible negative impacts from the approach (Hems & Flatau, 2011a, 2011b). This impacted on the research through the increased workload experienced by Burnside as they participated in the trial, meaning that they organisation was unable to participate in the research as extensive as intended. The other development of note was the acknowledgement by Fair Work Australia that gender was the primary factor in the undervaluation of social and community service workers practicing in non-government organisations (Fair Work Australia, 2012) such as those included in this research. The NSW state government and the Australian Government have since committed to these pay rises (ABC News, July 15, 2012,
July 16, 2012. While this change did not occur until December 2012, the indirect impact of an impending 23-45% pay rise is a significant event to occur over the course of the study. While no participants talked about this important change, it remains as part of the subtext of the research.

4.7 The Research Participants

While engaged with the research sites, my intent was to interview as many individuals as possible who had experience of the evaluation. The sampling of these individuals represents a combination of typical case sampling (Patton, 1990, p. 173) and intensity sampling (Patton, 1990, p. 171), as the participants represented for the most part the typical experience of the phenomenon, but by being identified by the organisations’ gate-keeper were more likely to be information rich cases with some significant experience of the phenomenon to share.

Following my case study protocol, in each case I started by working through my initial contact, who was the gatekeeper to conducting the research, as well as a high level informant of how the evaluation played out. I asked to interview two kinds of participants (Table 4.1), who would have the kinds of information I was looking for about the experience and perceptions of the evaluation. I asked for respondents who could speak to the way the evaluation had been managed and implemented, and could talk about the organisational impact it had. These kinds of respondents were able to talk about how the evaluation was conducted, and could put it into the organisational context. They were also able to talk to the steps that had been taken by the organisation to foster the influence of the evaluation, primarily through changes to policy and operational requirements. The second group of respondents I asked for access to were the front-line human service practitioners who were directly involved in working with service users. While initially I wanted to find practitioners that had
been around for all phases of the evaluations, I found that this was ambitious considering the normal turnover in human service organisations (Healy, et al., 2009). Some participants who were able to offer a comprehensive view of the evaluation, while others had only been around for part of the evaluation, but could speak to the impact the report had on practice.

Table 4.1 Description of Participant Groups

<table>
<thead>
<tr>
<th>Groups</th>
<th>Evaluation Data</th>
<th>Contextual Data</th>
<th>Position Titles</th>
</tr>
</thead>
<tbody>
<tr>
<td>Managers, Research and Evaluation Staff</td>
<td>What led to the evaluation being done; why was it done; how was the evaluation conducted; what has occurred in terms of working with the data and findings; what has been the impact of the evaluation;</td>
<td>The person’s role in the organisation and relation to the program; what the program does and how it fits into the organisation; relationship and interaction with the funder; what data collection is required by the funder; organisational commitment to evidence based practice; what information do practitioners need</td>
<td>Manager Research &amp; Evaluation, Site Manager of Program, Manager of Research and Program Development,</td>
</tr>
<tr>
<td>Service Level Staff</td>
<td>How the evaluation was undertaken; the impact of the evaluation from the practice level</td>
<td>The person’s role in the program; what the program does; prior experiences or understanding of evaluation; how evidence based practice plays out; what information do practitioners need</td>
<td>Family Support Coordinator, Caseworker, Team Leader, Community Connector, Child and Family Worker</td>
</tr>
</tbody>
</table>

4.7.1 The evaluation of Brighter Futures

The first case study is concerned with a program run by the Benevolent Society called Brighter Futures. Participants were suggested and contacted through the research and evaluation manager. As described in the previous chapter, the process
followed the case study protocol (Appendix A) of having the gatekeeper identify key respondents and obtain permission for me to invite them to participate in the study. The first group of participants were contacted on the 27th June 2011, and were given the opportunity to opt-out of being contacted by me about participating in the study. While I was initially cautious about the selection of participants by the research and evaluation manager, I accepted this as part of conducting organisational research; being an outsider I lacked the organisational knowledge to be able to identify the informants I needed. On reflection, my concerns about not having control of my sample seem unfounded, the research and evaluation manager demonstrated a genuine interest in exploring practice learning and how evaluation work was perceived in the organisation. In retrospect this is unsurprising considering the organisation was in the process of developing a strategic policy for evaluation. After allowing two weeks for the potential informants to opt-out, I contacted the suggested participants and emailed them an information sheet (Appendix B). Initially, I got a response from just two people, but in the course of later interviews, I found that all my participants were part of a group that had regular meetings about the program, and that the research and evaluation manager had encouraged them all to participate in my study. Over the course of the research all the people that had consented to be contacted ended up participating, meaning I had a comprehensive view of the program across all the sites where it was delivered. While it was a good outcome to have all of the members participate, some concerns remain about just how voluntary participation was in my study.

My second group of participants were much more disconnected from the central structure of the two organisations, and much more embedded in the community associated with their site. While I was hoping to get a majority of
participants in practice roles without team leadership or management responsibilities, knowledge of the evaluation seemed to be very low at that level. The two participants in purely practice roles had next to no knowledge of the evaluation taking place or its dissemination. All members of my second group of participants were directly involved in human service practice, but also had responsibility for the supervision of other practitioners or for the management of the site. In this sense, they were able to speak to not only the influence of the evaluation on their practice, but that of their staff, and of the discussions that went on that might relate to interpersonal mechanisms of influence.

As can be seen in Table 4.2, obtaining the interviews was a protracted process that took place over the course of a year. The constant challenge was in finding a convenient time for the participant to be interviewed. The following tables provide an overview of the data sources and timeline or relevance to this research.
Table 4.2 Brighter Futures Participant Details (Pseudonyms have been used to replace names)

<table>
<thead>
<tr>
<th>Name</th>
<th>Grouping</th>
<th>Location</th>
<th>Interview</th>
<th>Date of Interview</th>
<th>Length of Interview</th>
</tr>
</thead>
<tbody>
<tr>
<td>Victor</td>
<td>BF Managers Group</td>
<td>Metro</td>
<td>In-Person Interview</td>
<td>21/06/2011</td>
<td>50:51</td>
</tr>
<tr>
<td>Victor (Second Interview)</td>
<td>BF Managers Group</td>
<td>Metro</td>
<td>In-Person Interview</td>
<td>29/05/2012</td>
<td>21:31</td>
</tr>
<tr>
<td>Lauren</td>
<td>BF Managers Group</td>
<td>Regional</td>
<td>Phone Interview</td>
<td>19/07/2011</td>
<td>45:32</td>
</tr>
<tr>
<td>Earl</td>
<td>BF Managers Group</td>
<td>Metro</td>
<td>In-Person Interview</td>
<td>19/07/2011</td>
<td>50:12</td>
</tr>
<tr>
<td>Megan</td>
<td>BF Managers Group</td>
<td>Metro</td>
<td>In-Person Interview</td>
<td>26/07/2011</td>
<td>50:16</td>
</tr>
<tr>
<td>Vanessa</td>
<td>BF Managers Group</td>
<td>Metro</td>
<td>In-Person</td>
<td>04/08/2011</td>
<td>1:00:20</td>
</tr>
<tr>
<td>Courtney</td>
<td>BF Managers Group</td>
<td>Regional</td>
<td>Phone Interview</td>
<td>19/10/2011</td>
<td>37:58</td>
</tr>
<tr>
<td>Stephanie</td>
<td>Team Leader</td>
<td>Regional</td>
<td>Phone Interview</td>
<td>26/09/2011</td>
<td>43:41</td>
</tr>
<tr>
<td>Rebecca</td>
<td>Caseworker</td>
<td>Metro</td>
<td>In-Person</td>
<td>10/10/2011</td>
<td>44:25</td>
</tr>
<tr>
<td>Natalie</td>
<td>Team Leader</td>
<td>Metro</td>
<td>In-Person</td>
<td>26/10/2011</td>
<td>43:43</td>
</tr>
<tr>
<td>Monica</td>
<td>Team Leader</td>
<td>Metro</td>
<td>Phone Interview</td>
<td>19/04/2012</td>
<td>42:41</td>
</tr>
<tr>
<td>Christine</td>
<td>Family Worker</td>
<td>Metro</td>
<td>Phone Interview</td>
<td>31/04/2012</td>
<td>21:29</td>
</tr>
<tr>
<td>Miranda</td>
<td>Team Leader</td>
<td>Regional</td>
<td>Phone Interview</td>
<td>02/05/2012</td>
<td>15:49</td>
</tr>
</tbody>
</table>
Table 4.3 Brighter Futures Organisational Documents included in Analysis

<table>
<thead>
<tr>
<th>Document</th>
<th>Document Title</th>
<th>Date</th>
<th>Access</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Performance Improvement Plan Template</td>
<td>2012</td>
<td>Internal Document</td>
</tr>
<tr>
<td>B</td>
<td>Brighter Futures Early Intervention Program: Interim Report 1</td>
<td>2008</td>
<td>Public Record</td>
</tr>
<tr>
<td>C</td>
<td>Brighter Futures Early Intervention Program: Interim Evaluation Report</td>
<td>2009</td>
<td>Public Record</td>
</tr>
<tr>
<td>D</td>
<td>Brighter Futures Early Intervention Program: Final Report</td>
<td>2010</td>
<td>Public Record</td>
</tr>
<tr>
<td>F</td>
<td>Evaluating Evaluation Capacity Building in a Large NGO</td>
<td>2012</td>
<td>Public Record</td>
</tr>
<tr>
<td>G</td>
<td>External Influences Shaping the Evaluation of a Tertiary Child Protection Program</td>
<td>2010</td>
<td>Public Record</td>
</tr>
<tr>
<td>H</td>
<td>Making the Most of Evaluation: A Provider Perspective on the Outcomes of Brighter Futures</td>
<td>2010</td>
<td>Public Record</td>
</tr>
<tr>
<td>J</td>
<td>The Benevolent Society Organisational Structure</td>
<td>2008</td>
<td>Public Record</td>
</tr>
<tr>
<td>K</td>
<td>Evaluation Plan: Early Intervention Program</td>
<td>2006</td>
<td>Public Record</td>
</tr>
<tr>
<td>L</td>
<td>Policy &amp; Procedures Evaluation</td>
<td>2012</td>
<td>Internal Document</td>
</tr>
</tbody>
</table>

Table 4.4 Brighter Futures Timeline of Significant Events

<table>
<thead>
<tr>
<th>Event</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>Provision of Funding to Community Services for Early Intervention</td>
<td>Late 2002</td>
</tr>
<tr>
<td>Evaluation Begins</td>
<td>2006</td>
</tr>
<tr>
<td>Data Collection Begins</td>
<td>2007</td>
</tr>
<tr>
<td>First Report</td>
<td>May 2008</td>
</tr>
<tr>
<td>Second Report</td>
<td>July 2009</td>
</tr>
<tr>
<td>Data Collection Concludes</td>
<td>2009</td>
</tr>
<tr>
<td>Final Report</td>
<td>April 2010</td>
</tr>
<tr>
<td>Case Study Interviews Begin</td>
<td>June 2011</td>
</tr>
<tr>
<td>Implementation of Revised Brighter Futures Program Guidelines</td>
<td>January 2012</td>
</tr>
<tr>
<td>Case Study Interviews Conclude</td>
<td>May 2012</td>
</tr>
</tbody>
</table>

4.7.2 The evaluation of the St Marys Child and Family Network
The process of getting access to participants in the program managed by Uniting-Care Burnside occurred in a similar way to the previous case, through discussion and negotiation with the manager of research in the organisation. Based on a discussion about the evaluation and some of the documents I had been given, I selected the St
Marys Child and Family Network evaluation and obtained organisational permission to undertake the interviews and access relevant documents.

All the practitioner interviews were done over the course of two days in the programs room at the St Marys centre. The participants comprised all of the workers in the centre at the time, as they all consented to participating in the research. Many of the participants were quite confused about what the research was for, many of them assumed that I was from the university that had done the evaluation and what I was doing was an official evaluation of the evaluation. This plus the fact that I had come in through ‘head-office’ meant that I had considerable access and universal participation at the centre.

Almost a year later, while waiting for the evaluation to be disseminated I undertook an interview with the manager of research in Burnside on the progress of the evaluation and the program, and on the organisational view of what had happened with the evaluation.

*Table 4.5 Burnside Participant Details (Pseudonyms have been used)*

<table>
<thead>
<tr>
<th>Name</th>
<th>Grouping</th>
<th>Location</th>
<th>Interview</th>
<th>Date of Interview</th>
<th>Length of Interview</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lauri</td>
<td>Manager</td>
<td>Metro</td>
<td>In-Person</td>
<td>10/05/2012</td>
<td>1:03:14</td>
</tr>
<tr>
<td>Jessica</td>
<td>St Marys Team Member</td>
<td>Metro</td>
<td>In-Person</td>
<td>09/07/2011</td>
<td>39:45</td>
</tr>
<tr>
<td>Nicky</td>
<td>St Marys Team Member</td>
<td>Metro</td>
<td>In-Person</td>
<td>09/07/2011</td>
<td>42:03</td>
</tr>
<tr>
<td>Amy</td>
<td>St Marys Team Member</td>
<td>Metro</td>
<td>In-Person</td>
<td>09/07/2011</td>
<td>54:28</td>
</tr>
<tr>
<td>Claudia</td>
<td>St Marys Team Member</td>
<td>Metro</td>
<td>In-Person</td>
<td>10/07/2011</td>
<td>39:29</td>
</tr>
<tr>
<td>Susan</td>
<td>St Marys Team Member</td>
<td>Metro</td>
<td>In-Person</td>
<td>10/07/2011</td>
<td>21:25</td>
</tr>
<tr>
<td>Zoe</td>
<td>St Marys Team Member</td>
<td>Metro</td>
<td>In-Person</td>
<td>10/07/2011</td>
<td>1:08:10</td>
</tr>
</tbody>
</table>
### Table 4.6 Burnside Organisational Documents included in Analysis

<table>
<thead>
<tr>
<th>Document</th>
<th>Document Title</th>
<th>Date</th>
<th>Access</th>
</tr>
</thead>
<tbody>
<tr>
<td>B</td>
<td>A Collaborative Research Project: Exploring the Efficacy of a Newly Established Integrated Service for Children, Families and Service Staff</td>
<td>2010</td>
<td>Public Record</td>
</tr>
</tbody>
</table>

### Table 4.7 Burnside Timeline of Significant Events

<table>
<thead>
<tr>
<th>Event</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>Evaluation Data Collection Began</td>
<td>August 2008</td>
</tr>
<tr>
<td>Case Study Interviews Undertaken</td>
<td>July 2011</td>
</tr>
<tr>
<td>Research Manager Interview Undertaken</td>
<td>May 2012</td>
</tr>
</tbody>
</table>

#### 4.8 Data Collection

As discussed in the methodology section, this research is constituted by a phase of case studies in order to provide a context for the narrative of evaluation influence using the participants’ accounts, followed by an analysis of the factors associated with this influence. While the data used for these two stages overlap, I have described the data collection for these stages separately in order to distinguish the intended purpose.

#### 4.8.1 Case Study data

Following the process of obtaining permission from the organisation, the research began with case studies in order to get a sense of the organisations, the program, and the environment staff work in. In order to provide some scope for the case study, the data collection was directed towards answering two main questions about the research context:

- What are the goals of the program, and how are the resources and staff organised towards pursuit of these goals?
• What was the narrative of the evaluation from the perspective of the evaluators (i.e. what were the course of events that lead to and followed on from the evaluation occurring)?

This was undertaken through an analysis of a number of documents, primarily the evaluation reports, tender documents, documents describing policy or practice guidelines, and dissemination reports. As mentioned a case study protocol (Appendix A) was used in order to provide some direction and consistency to the process of investigating the organisations; this process was followed until enough information was obtained to adequately describe each case. As the interviews with participants were embedded within the case studies, these are also detailed in the protocol.

4.8.2 Phenomenological data
In-depth semi-structured interviews with staff working within the organisation provided the second data set. It was particularly important that the interviews were recorded, as the chosen method of the present research involves going beyond the descriptions of the participant; any interpretations of the data need to be based on a faithful representation of the interview (Wengraf, 2001).

The methodology outlined for the present research emphasises the importance of co-construction, in practice this requires some latitude for the participant to lead the interview (Bryman & Bell, 2007), with the researcher’s responsibility to provide some structure and purpose to the dialogue. While the interview schedule (Appendix C) is quite lengthy and structured, effort was made during the interviews to tailor the questions to make them more relevant to the participant and to ask the questions in a conversational way. The schedule is composed of a number of broad overarching questions with probes to direct the participant to addressing specific sub-questions. If in the course of the conversation initiated by the main questions, sub-questions were not addressed, the probes were used to get a more specific response to part of the
The use of probes and other clarificative devices in the interview fit within the view that phenomenological interviews are characterised by “use of reflection, clarification, request for examples and description and the conveyance of interest through listening techniques” (Jasper, 1994, p. 311). From a broad list of topic areas and sample questions, the schedule was developed through a number of pilot interviews, which assisted in tailoring the questions to more closely reflect the experiences of the participants, improving their engagement and participation in the research. The interview schedule sought to elicit description of the phenomena, but also the broader experiences and perceptions of the participants in order to develop an understanding of “… the individual’s being-in-the-world and how these meanings influence the choices they make” (Lopez & Willis, 2004, p. 729).

Wherever possible interviews were undertaken in person, with the remainder (mainly with respondents from regional NSW) undertaken by phone. Conscious of the effect that setting can have on a participant’s responses, and of wanting to minimise any inconvenience to the participants, the interviews took place at the participant’s leisure and at a setting of their choosing. For the Brighter Futures evaluation, interviews were arranged individually, and took place over a twelve month span of time. For the St Marys evaluation, two days were arranged for me to come down to the site and interview service staff individually. Following an extended period of time I then followed up and interviewed a manager level respondent. Considering the orientation towards co-construction explicit in the research approach, special attention was paid toward establishing trust, rapport, and to fostering exchange and discussion on the research topic. The recording of the interview began after obtaining written consent and a brief discussion of the research aims in order to orient and set the participant at ease.
4.9 Data Analysis Procedures

As is common with research through qualitative interviewing, there is some overlap between the process of data collection and data analysis. During the course of the interview, notes were taken to record initial reflections. Notes were also taken directly after the interview. These notes then complimented the partial transcription of the recorded interview. While initially I had planned to transcribe my own interviews, and to attempt to do this within 48 hours of the interview occurring, this proved to be onerous and not an effective use of my time. I undertook a verbatim transcription of seven interviews myself, hired a professional research transcription service for six interviews that were particularly rich in data, and selectively transcribed the rest of the interviews (seven). This process involved listening through the interviews and transcribing only that information relevant to the research area.

Figure 4.2 shows the data drawn on for each of the data analysis stages. While the data collection occurred as a single stage, the analysis is separated into three phases consistent with the aims of the research:

- Obtain a comprehensive understanding of two evaluations, which was pursued through a simple case study (Stake, 2003);
- Use Mark and Henry’s (2004) evaluation influence framework to identify instances where the evaluation has affected change; and
- From the evaluation influence mechanisms identified, conduct a phenomenological analysis of the experiences of participants in order to infer a set of conditions and barriers for practitioner learning from evaluation, and consider the implication of these for evaluation practice.
4.9.1 Case Studies

As described above, the approach to the case studies represents what Stake (2003) calls a simple case study. The documents that described the evaluation were reviewed and analysed for the narrative they produced of how the evaluators described the evaluation. Annotated notes, quotes and references to the original documents were categorised in a number of headings. These headings present descriptive detail about (a) the organisation the program is based in, (b) the program itself, what its intended outcomes are and how it is delivered, (c) the evaluation of the program, including the methodology of the evaluation, the main findings, and how it was disseminated.

The participant interviews often provided additional context and understanding in the preparation of the case studies; illustrating the events detailed in the documents. As the interviews were undertaken concurrently with the case study, the researcher’s understanding of the cases developed with the participants’
perspectives of the events of the evaluation in mind. The participants highlighted many issues that in addition to forming part of the analysis in phase three prompted further exploration in the documents for the case study.

The second analytic approach was to apply the Mark and Henry (2004) framework of evaluation influence to these accounts and the participants’ accounts of how the evaluation has influenced change in the organisation or the program. While the focus is on practitioner learning from evaluation, all types of influence were included in the analysis in order to be able to investigate their connection to learning. The next section provides some detail about this analysis process and the definition used for the different influence mechanisms.

4.9.2 The identification of mechanisms of evaluation influence

Evaluation influence has been reviewed extensively in chapter two, focusing on the use of influence frameworks to identify mechanisms and interconnected threads of influence. Despite a developing body of literature, there is relatively little guidance on how to define each of these mechanisms, and how to demonstrate a link between mechanisms retrospectively. This remains a challenge, particularly in using evaluation influence as an analytical framework in case study research. I include a description of the influence mechanisms here in the methodology chapter, as Mark and Henry’s (2004) framework is key to understanding the analysis phase of the research.

The following emphasises the description of the mechanisms most relevant to practitioner learning, namely the individual and interpersonal, as many of the collective mechanisms are self-explanatory. Two recent conceptual additions to Mark and Henry’s (2004) framework should also be mentioned, particularly as they relate to the relationship between organisations. Oliver (2008) suggests an additional level of influence called intra-organisational, that mirrors the collective level in large
multi-site organisations, with a focus on policy learning and change. This proposed level is not quite as relevant at the scale of the organisations included in this research, because while there were differences in practices between sites in Brighter Futures, they did not have the kind of policy and procedural autonomy described in Oliver (2008). Appleton-Dyer, Clinton, Carswell, & McNeill (2012) add the complexity of public sector partnerships to the model, suggesting that evaluation influence is mediated through the characteristics, functioning, and evaluation behaviour related to partnerships. While the nature of the partnership with Community Services was considered in the analysis, the description of how the partnership operated was purely from the perspective of staff within The Benevolent Society.

4.9.2.1 Individual mechanisms

Mark and Henry (2004) describe individual mechanisms as the starting point of change, suggesting that collective and interpersonal change begins from the internalised change processes that occur as a result of an evaluation. Table 4.8 outlines these mechanisms, and the rest of the section will present some detail on how they have been defined by Henry and Mark (2003), Mark and Henry (2004), and other researchers (Fjellstrom, 2007; Fleming, 2011; Oliver, 2008; C. H. Weiss, Murphy-Graham, Petrosino, & Gandhi, 2008).
Elaboration is the starting point of individual change. Mark and Henry (2004) describe it as an attempt at a systematic cognitive process incorporating the new information. They suggest this can be measured by the amount of time and effort spent thinking in response to evaluation information. Fleming (2011) links elaboration to a thoughtful process of attitude change, where “individuals think carefully about issue relevant information presented in a persuasive message, and they base their attitude of their idiosyncratic cognitive responses to the message” (p. 215). In this sense elaboration represents a pre-cursor to individual learning, an openness and receptiveness to drawing on information about an issue. Oliver (2008) suggests that in retrospective case studies, elaboration may be difficult to discover, however a number of studies have reported finding elaboration as a discrete outcome of evaluation (Fjellstrom, 2007; Weiss, et al., 2008).

The mechanism Mark and Henry (2004) call heuristics represents a short cut to learning, where undertaking an exhaustive search for information is bypassed in

<table>
<thead>
<tr>
<th>Level of Analysis</th>
<th>Mechanisms</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>General</td>
<td>Elaboration</td>
<td>Thinking about an issue in response to evaluation information.</td>
</tr>
<tr>
<td></td>
<td>Heuristics</td>
<td>Formation or change of attitudes without thinking comprehensively about information. Also reliance on quick answers from experiential learning and intuition.</td>
</tr>
<tr>
<td></td>
<td>Priming</td>
<td>Makes an idea or concept central to the judgement of an issue.</td>
</tr>
<tr>
<td></td>
<td>Skill Acquisition</td>
<td>Development of skills from participation in an evaluation.</td>
</tr>
<tr>
<td>Cognitive/Affective</td>
<td>Salience</td>
<td>Increasing the importance of an issue to an individual; a priority shift.</td>
</tr>
<tr>
<td></td>
<td>Opinion/Attitude Valence</td>
<td>Change in the beliefs and attitudes about a program or the underlying assumptions of the program through evaluation.</td>
</tr>
<tr>
<td>Motivational</td>
<td>Personal Goals and Aspirations</td>
<td>Change to the goals a person aims to achieve in their work related to evaluation.</td>
</tr>
<tr>
<td>Behavioural</td>
<td>New Skill Performance</td>
<td>Skill acquisition put into practice.</td>
</tr>
<tr>
<td></td>
<td>Individual Change in Practice</td>
<td>Thoughtful change to performance based on reflection.</td>
</tr>
</tbody>
</table>
favour of searching for answers through experiential learning, experimentation, and intuition (Hogarth, 1981). Fleming (2011) links heuristics as an evaluation influence mechanism to a kind of non-thoughtful process of attitude change. The heuristic-systematic model (Chaiken, 1980, 1987; Chaiken, Liberman, & Eagly, 1989) contrasts systematic thinking and heuristic thinking; systematic thinking reflects individuals thinking comprehensively about the information. Heuristic thinking involves the formation or change of attitudes through heuristics such as “experts can be trusted, majority opinion is correct, and long messages are valid messages” (Chaiken & Maheswaran, 1994, p. 460). In the context of evaluation, heuristics may involve individuals searching for cues to help them decide whether to change their attitudes without having to reflect on and integrate the evaluation information into their experiences and values. In short heuristics bypasses the learning processes discussed by Antonacalopau (2001, 2006, 2008), to produce a non-thoughtful attitude change that tends to be ineffective at influencing others (Petty, Barden, & Wheeler, 2009; Priester, Nayankankuppam, Fleming, & Godek, 2004).

Much like elaboration, priming represents a cognitive process that is likely to be earlier in the chain of influence. Priming concerns the effect an evaluation can have in highlighting a particular idea or concept in a way that affects the way a broader issue is understood or thought about (Henry & Mark, 2003). In other words this means that priming influences judgements, attitudes, and actions through making an idea or concept central to the judgement of an issue, or reducing the importance of others (Oliver, 2008). In operationally defining priming for her case study, Oliver (2008) describes it as an issue rising to the forefront due to the way information is presented or because of the circumstance of an event. Priming represents a
mechanism that may affect later learning processes, the evaluation may raise issues that become central in the practitioner’s thinking about the program.

Henry and Mark (2003) and Oliver (2008) describe skill acquisition as the development of new abilities primarily through participation in an evaluation. This may include evaluative thinking skills, the ability to work collaboratively, or to deliberate with others about evaluation findings. Skill acquisition reflects direct and individual learning, primarily coming from the process of the evaluation. In Diaz-Puente, Yague and Afonso (2008) one of the objectives of the evaluation was to develop evaluation capacity within the organisation. Through training and peer learning, individuals developed and practiced evaluation skills (the collection, analysis, and interpretation of project data). The acquisition of knowledge and skills about evaluation may be a powerful early process in changing attitudes and behaviours at an individual level, however this may depend on the quality of the evaluation data among other factors (Preskill & Boyle, 2008).

Henry and Mark (2003) define the cognitive/affective mechanism of salience as “... the importance of an issue as judged by an individual” (p. 300). Influence through this mechanism reflects the effect an evaluation can have in increasing the importance of a particular issue to an individual. Similarly, Oliver (2008) defines it as beyond a change in attitude, representing a priority shift. Henry and Mark (2003) provide the example of Fuller, Kagan, Caspary, and Gauthier’s (2002) evaluation of the effects of welfare reform on children in raising the long term salience of children’s wellbeing when individuals thought about welfare reform. While McEathron (2008) suggests this mechanism is primarily a quality of the evaluation information, salience depends on a general process to lead to this priority shift.
The general mechanisms discussed above all link into opinion/attitude valence, a vital mechanism linking preliminary cognitive processing about an issue to more significant changes in values and eventually behaviour. It concerns the changing of beliefs and attitudes about a program or the underlying assumptions the program is built upon through evaluation (Henry & Mark, 2003; Oliver, 2008). Along with the kind of thoughtful processing discussed above, attitude change is more influential when individuals do not have strong pre-existing attitudes (Fleming, 2011). Attitude change primarily concerns some aspect of the program, however it also can describe changes in attitudes about evaluation. In relation to learning, attitude change represents the outcome of an individual’s reflection process drawing on existing attitudes and values.

Motivational processes conceptualise influence in relation to goals and aspirations, and the perception of rewards and punishments in one’s daily work (Mark & Henry, 2004). While other groups of processes concern thoughts or actions, motivational processes relate to perceptions of what is important, and the incentives at play. At an individual level evaluation can influence personal goals by affecting the personal goals and aspirations important to individuals in their work (Mark & Henry, 2004).

Henry and Mark (2003) describe individual behaviour change as a mechanism of influence, calling it “direct instrumental use” (p. 302). Oliver (2008) describes it as a change in the way an individual acts as a direct result of participation or reading an evaluation report. Behaviour change will likely represent the long term outcome of influence, leading from any number of other mechanisms (Mark & Henry, 2004), but equally behaviour change at any level will likely continue to link to additional mechanisms. Mark and Henry (2004) describe
behaviour change as either new skill performance or individual change in practice.

*New skill performance* represents behavioural change without the normative baggage of replacing existing values and norms. In terms of the kind of reflective process of learning discussed, the development of new skills requires a sense from the participants that these are skills worth having (Preskill & Boyle, 2008). The behavioural mechanism of *individual change in practice* reflects a change to existing processes and practice. This change may come about from individual level processes or indirectly through collective and interpersonal level processes, however the end result is an individual making a thoughtful change to their performance based on their reflection (Fleming, 2011).

**4.9.2.2 Interpersonal mechanisms**

Interpersonal mechanisms act as an intermediary form of influence that is used to affect a change in the attitudes and behaviours of others. Evaluation serves as a kind of ammunition for discussion, argument, and debate, helping to support particular points of view, but also can be used in reframing the discussion around a topic. Valovirta (2002) highlights the importance of these types of mechanisms by suggesting that the use of evaluation information effectively amounts to a process of argumentation and persuasion. From previous research using retrospective case studies, it can be said that there are inherent challenges in identifying some interpersonal mechanisms, particularly interactions the researcher was not present for (Weiss, et al., 2005). As an example, Oliver (2008) discussing an evaluation that had concluded before his research found no specific instances of interpersonal influence, yet interviewees alluded to interpersonal dynamics and the effect individuals had in creating broader change. However, studies concerned with examining influence at a policy level (Lehtonen, 2010), and studies that began alongside the evaluation (Diaz-
Puente, et al., 2009; Diaz-Puente, et al., 2008; Fjellstrom, 2007) were much more successful at finding examples of interpersonal influence.

Table 4.9 Definitions of the Interpersonal Mechanism of Evaluation Influence

<table>
<thead>
<tr>
<th>Level of Analysis</th>
<th>Interpersonal</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>General</td>
<td>Justification</td>
<td>Evaluation findings used to support an existing position in attempting to influence others.</td>
</tr>
<tr>
<td></td>
<td>Persuasion</td>
<td>Drawing on information from the evaluation to change the attitudes of others.</td>
</tr>
<tr>
<td></td>
<td>Change Agent</td>
<td>A person who actively pursues change as a result of some change in awareness from the evaluation.</td>
</tr>
<tr>
<td>Minority-Opinion Influence</td>
<td></td>
<td>Individuals with a minority opinion bring about change through evaluation information.</td>
</tr>
<tr>
<td>Cognitive/Affective</td>
<td>Local Descriptive Norms</td>
<td>Change to the attitudes and values at the local level.</td>
</tr>
<tr>
<td>Motivational</td>
<td>Injunctive Norms</td>
<td>Change to the shared standards of appropriate behaviour (driven by reward and punishment) influenced by the evaluation.</td>
</tr>
<tr>
<td></td>
<td>Social Reward</td>
<td>Social inducement to adopt a particular behaviour related to the evaluation.</td>
</tr>
<tr>
<td></td>
<td>Exchange</td>
<td>Actual inducement to adopt a behaviour.</td>
</tr>
<tr>
<td>Behavioural</td>
<td>Collaborative Change in Practice</td>
<td>Change in practices as part of an interaction or persuasive process.</td>
</tr>
</tbody>
</table>

Much like symbolic use, justification refers to the use of evaluation findings to support a previously held position in attempting to influence others (Henry & Mark, 2003; Oliver, 2008). Much as this has been disparaged, justification has an important place in influencing others and collectives towards a particular course of action, and in developing a robust arguments for viewpoints. Henry and Mark (2003) accept the role evaluation has to play in justifying existing viewpoints as long as the use is not inaccurate or selective in using the information. However there are challenges particularly in using more participatory and equivocal evaluations in this way, where the interpretation of the evaluation can lend support to multiple points of view.

Justification as an influence mechanism represents a point before persuasion, where the evaluation is used to support the position or simply to make the case for the
importance of the issue. Fleming (2011) highlights that a well thought out viewpoint is likely to be more influential and persuasive down the chain of influence. Justification serves as a starting point to some of the other mechanisms of interpersonal change.

A further mechanism that will be sought out in the data is that of persuasion, this involves communication that draws on information from the evaluation to change the attitudes of others (Henry & Mark, 2003), primarily evaluation findings (Oliver, 2008). The evaluation literature reflects the fact that evaluation is rarely directly used, that is to say persuasion is not a direct process, it involves dialogue, discussion and argumentation (Valovirta, 2002). Persuasion reflects the development of arguments that are thoughtful and convincing to others who hold different viewpoints, which involves different processes to those who hold the same view or have no strong view (Cialdini, 2009; Petty & Cacioppo, 1986). The interaction and communication that derives from an evaluation, both from parties directly involved in the evaluation and those merely affected, forms part of the social production of knowledge (Suarez-Herrera, Springett, & Kagan, 2009), and adds to the capacity of the organisation to use and understand evaluation information (Forss, Rebien, & Carlsson, 2002).

Where the process or findings of an evaluation lead to an individual focusing on changing how things are done within the organisation, Henry and Mark (2003) refer to this as a change agent. Oliver (2008) also suggests that groups can act as change agents, but more to the point a change agent is one who actively pursues change as a result of the evaluation. Fleming (2011) provides the example of the dean of a university acting as a change agent, going from reading the evaluation
evidence supporting senior comprehensive exams, and changing her attitude, to becoming an active and persuasive advocate of the exams.

Minority-opinion influence refers to situations where individuals or groups holding minority views can bring about change through the marshalling of evaluation information (Henry & Mark, 2003). Oliver (2008) suggests this is a similar mechanism to persuasion, but through a minority opinion group. Minority-opinion groups can be particularly effective in challenging implicit norms with comprehensive, consistent and well-thought out arguments (Alvaro & Crano, 1996; Wood, Lundgren, Ouellette, & Busceme, 1994). Henry and Mark (2003) provide the example of the lobby group Mothers Against Drink Driving (MADD) who were able to influence change through their persuasive use of information and moral position.

Cognitive/affective mechanisms at the interpersonal level are described in Mark and Henry (2004) as local descriptive norms, which Oliver (2008) defines as where evaluation affects the social behaviour of stakeholders. This can be thought of as the influence of a change of collective attitudes and culture on others.

In terms of motivational processes Henry and Mark (2003) describe injunctive social norms as “shared standards about what is an appropriate or acceptable behaviour in a specific context and at a particular time” (p. 302). Oliver (2008) describes it in terms of evaluation affecting the agreed upon principles about how to conduct oneself in a given setting. Evaluation can serve to inform, challenge, or disseminate social norms, producing an effect on an individual via group norms.

Whereas local descriptive norms is concerned with attitudes and values, social norms is based on human responses to rewards and punishments, goals and aspirations (Mark & Henry, 2004). Also included as mechanisms under motivational processes are social reward and exchange; both describing a kind of reinforcement for
behaviour. Neither of these mechanisms have received any real definition beyond their names in the existing literature, but are straightforward in the sense that motivations are influenced interpersonally by some kind of social or material inducement.

Mark and Henry (2004) describe behavioural change at the interpersonal level as *collaborative change in practice*, which encapsulates the effect of the general, affective, and motivational influences. Distinct from individual changes in practice, or skill performance, this mechanism describes a persuasive message having the effect of changing behaviour as part of an interaction or argumentation (Valovirta, 2002).

**4.9.2.3 Collective mechanisms**

As many of these are fairly self-explanatory, these mechanisms are briefly described in Table 4.10.
Table 4.10 Collective Mechanisms of Evaluation Influence (Adapted from Mark & Henry, 2004, p. 41)

<table>
<thead>
<tr>
<th>Collective</th>
<th>Collective</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level of Analysis</td>
<td>Mechanisms</td>
</tr>
<tr>
<td>General</td>
<td>Ritualism</td>
</tr>
<tr>
<td>Legislative Hearings</td>
<td>Evaluation is drawn on in the discussion and proposal of legislation.</td>
</tr>
<tr>
<td>Coalition Forming</td>
<td>The formation and alignment of groups sharing a position on an issue related to the evaluation.</td>
</tr>
<tr>
<td>Drafting Legislation</td>
<td>Evaluation is drawn on in the drafting of legislation.</td>
</tr>
<tr>
<td>Standard Setting</td>
<td>Evaluation results in the setting or changing of standards.</td>
</tr>
<tr>
<td>Policy Consideration</td>
<td>Evaluation information enters into consideration in the policy making process.</td>
</tr>
<tr>
<td>Cognitive / Affective</td>
<td>Agenda Setting</td>
</tr>
<tr>
<td>Policy-Oriented Learning</td>
<td>Evaluation leads to an increased understanding at the policy-making level of the organisation.</td>
</tr>
<tr>
<td>Motivational</td>
<td>Structural Incentives</td>
</tr>
<tr>
<td>Market Forces</td>
<td>Incentives related to the functioning of markets change motivations related to some issue from the evaluation.</td>
</tr>
<tr>
<td>Behavioural</td>
<td>Program Continuation, Cessation or Change</td>
</tr>
<tr>
<td>Policy Change</td>
<td>Similar to policy-oriented learning, but resulting in policy change.</td>
</tr>
<tr>
<td>Diffusion</td>
<td>Change resulting from a different evaluation.</td>
</tr>
</tbody>
</table>

4.9.3 Phenomenological analysis
So far section 4.9 has detailed the process of analysis for the case studies, and for the analysis of evaluation influence in each of the cases. This section presents the process of analysis for identifying the conditions and barriers to practitioner learning drawing on the identification of influence mechanisms. Each of the influence events identified serve as a probe into whether or not influence played out in that particular context, to add to the broader account of the participants on why the evaluation was influential. Focusing on the experiences and perceptions of participants associated with these influence events, this phase of the analysis develops an explanation of the observed regularities in the data (Blaikie, 2007). The result of this analysis is the presentation of a set of factors that based on the two case studies seems to explain the
factors associated with practitioner learning from evaluation, and the implications of these factors for evaluation practice.

The analysis included a four phase data analysis process guided by principles of hermeneutic interpretation, or a kind of cyclical flow of analysis of parts and the whole and the reinterpretation of the interpretations originally arrived at (Annells, 1996).

The first phase involved a line by line analysis of each of the transcripts and notes; each transcript was read through several times with annotations made and key phrases highlighted. This process was guided by Van Manen’s (1990) selective reading approach, which asks of the researcher “what statements of phrases seem particularly essential or revealing about the phenomenon or experience being described” (p. 93), particularly in parts of the interview where there was a sense that there was some important meaning below the awareness of the participant (Smith & Osborn, 2003). Attention was also given to statements that conveyed a taken for granted, or implicit collective belief or practice, and statements that contained ambiguity or “modalities and fluctuations” (Conroy, 2003, p. 3). As language is not only an important representation of the lifeworld of the participant, but often a reflection of institutionalised meaning and taken for granted assumptions, commonly used terms were noted order to get a sense of the language and meanings, and how they reflect real life experiences (Crowe, 2006).

The second phase of the analysis involved turning the sections of statements from the narratives into extracts according to provisional themes that emerged from the data (Klemm, 1986). These themes were arrived at through a reiterative process of organising the data around the issues the participants raised, continually revising and reflecting on the themes, allowing for insights that provided a better fit to the
data to prevail (Koch, 1999). Themes were highlighted in the transcription with exemplars to illustrate the basis for the categorisation (Cohen, Kahn, & Steeves, 2000). This process of spontaneously thinking through and interpreting the meaning in the accounts is referred to as an “experience of thinking” (Smythe, Ironside, Sims, Swenson, & Spence, 2008).

The third phase involved incorporating the information collected that allows some interpretation of the context the accounts can be understood as existing within. The holistic meanings of each narrative were reflected on in the context of the organisation and its values around evaluation, the program and its functioning, and how the evaluation was carried out.

The final phase of analysis focused on broad themes and categories; to step back from the microanalysis of themes and transcripts to be able to posit prominent experiential themes that unify and give sense to the stories. This final phase is the product of all the previous phases and reiterative processes combined to produce themes that link to and summarise the experiences and perceptions of the participants. This final stage also required the consideration of how the different categories correspond and interact, and the comparison of these themes to the existing literature to build theory.

4.10 Verification and Rigour

A number of strategies were employed to enhance the trustworthiness of the research, trustworthiness being constituted by credibility, dependability, confirmability, and transferability (Guba & Lincoln, 1989). A number of steps have been taken to enhance the credibility of the research: prolonged engagement in the form of the case studies of the research context in order to get a sense of the phenomena over time, and member checks of the transcripts and interpretations of
the themes and meanings of the individual interviews. Enhancing the dependability of the research requires attention to the quality of processes in data collection, data analysis, theory generation, and the stability of the data over time (Guba & Lincoln, 2005). The present research has attempted to demonstrate dependability by keeping extensive records and details of the thought processes that led to particular interpretive constructions and how these changed over time and in the light of new data, theory, or perspectives. Dependability is also promoted by incorporating reflexive processes into the interpretation of data. The relationship between these interpretations and the supporting data was the key focus of these records. Confirmability requires that the results and interpretations of the data are clear and logical inferences that stand up to the reader’s scrutiny; demonstrating confirmability requires the inclusion of relatively large extracts of raw data, allowing the reader the chance to examine for themselves the reasonableness of the researcher’s inferences. Finally, transferability depends on the research context, methods, and interpretations being described in such sufficient detail to enable the reader to judge the relevance of findings in different settings. Source triangulation will also serve as a further means of enhancing the trustworthiness of the research by the collection of information from documents, as well as participants in different parts of the case study organisations. While the research concerns the contextual interpretations of the participants, the collection of a number of different accounts from the same organisational context, and the collection of data concerning the context through the case studies will allow for the crosschecking and comparison of accounts.

4.11 Limitations of the Methodology

This research aimed to develop or extend theory, not test it. That said there are a number of limitations to this research that need to be acknowledged. While every
effort was made to find evaluations of social programs that aimed to foster practice learning, this proved to be quite difficult. Even though the final report suggested that identifying areas for change to improve service delivery was part of the evaluation, the Brighter Futures evaluation had a limited focus on program improvement and practice learning. Looking back to previous evaluation reports, the intention to identify areas for improvement did not appear until the final evaluation report. Although the intent to foster learning was limited in terms of the evaluation itself, The Benevolent Society took responsibility for disseminating and fostering the influence of the evaluation amongst practitioners. The Uniting-Care Burnside evaluation also had a limited focus on learning at the practice level. Although the evaluation was formative in wanting to understand how the integrated network was functioning, the evaluation had limited implications for day-to-day practice. This is a limitation in the sense that there were few examples of practitioner learning from evaluation to develop theory, and these examples were in a general environment that may not have been ideal for the phenomena to occur. The intent was to develop theory from looking at cases with the ideal conditions for learning; that said what was included perhaps represent fairly typical external evaluations of NGO provided human services.

The St Marys case study was not completed to the same extent as the Brighter Futures case study; it ended before the release of the final report and the dissemination of findings. As acknowledged, this case presented an opportunity to explore practitioner experiences and perceptions at the mid-point of an evaluation, but yet it was limited in the types of influence that could be observed if the case study had of been completed to the same extent as the Brighter Futures evaluation.
4.12 Ethical Considerations

The University of Western Sydney’s Human Ethics Research Committee (H8608) approved this research on the 1st of November 2010. Research proposals are examined based on the principles and themes of justice, beneficence, respect, risk and benefit, and participant’s consent (National Health and Medical Research Council, 2007). This research reflects these ideals. In undertaking research within organisations and within workplaces there are significant ethical issues to be addressed. The responsibility of the researcher to the participating organisations involved: minimising disruption to the work environment, providing an environment for the free consent of staff to either participate or not participate without experiencing significant consequences, and the allowing the organisation to decide if they were willing to be named. For all participants interviewed in person an information sheet and consent form was provided for them to sign (Appendix D). Participants who were interviewed by phone were read a consent statement paraphrased from the consent form.

A number of important ethical issues were considered over the course of the research. First, the researcher aimed to be clear to the participants about what the research is about, what the data would be used for. Secondly, with clear knowledge of what is involved in the research and what will be done with the information, written permission from participants was obtained. Thirdly, the confidentiality and anonymity of participants was maintained. While for the sake of description and clarity the aim was to include job descriptions and titles to provide some context of the participants’ role in the organisation, this proved impossible while providing anonymity. Fourthly, transcripts, analysis, and reports were made available to
participants at their request and participants retained the right to withdraw the inclusion of their transcript in the research until the final report is published.

For the two case studies, the research had to be approved by internal research ethics committees within each of the organisations. As well as the ethical treatment of participants, these committees considered the possible risks of the research in terms of the reputation of the organisations. In particular, naming the organisations was a key issue raised by these committees. Each of the organisations wanted to approve the final copy before the publication of the thesis, a stipulation that complicated some of the writing and editing processes of the thesis. In all published work prior to the completion of the thesis (Herbert, 2012) I have not named either of the case study organisations or the programs being delivered.

Participants may provide information that results in harm to themselves, others, or to the program itself. This was managed on a number of levels; member checking so that the participant agrees with the information and interpretation produced through interviews and retains a right to veto information they provided; triangulation through the analysis of organisational documents and interviews with service managers to improve the veracity of accounts; the maintenance of confidentiality and anonymity for individual participants.

The transcripts will be securely stored in a secured file for five years. After this point it will be reviewed if this information needs to be retained for future reference. Five years after completion represents the minimum period required under the National Statement on Ethical Conduct in Human Research. After this point, it can be reviewed if the information would be of use for future reviews on the subject.
4.13 Chapter Summary

This chapter presents a detailed methodology for the research, covering the criteria for methodological decision-making, the paradigms of inquiry inherent in the approach, the strategy and method, and the procedures undertaken. The chapter lays out the reasoning behind decisions around the research method, namely methodological pragmatism. Based on the research problem, a qualitative approach is presented as being most appropriate. The assumptions of qualitative research and criteria for reliability and validity of this type of research were presented.

The philosophical approach of the research is detailed in this chapter. Drawing on a depth realism ontology (Bhaskar, 1978), and moderate social constructionism (Hess, 1997; Van Fraassen, 1980), this thesis employs an abductive research strategy in order to describe meanings and interpretations of participants’ knowledge, to then develop social science concepts and theory (Bhaskar, 1979, 1986). In practice this means conducting case studies that present the experiences and perceptions participants have of evaluations intended to foster learning, followed by the development of theory and the application of existing theory about the factors involved in learning from evaluation.

Two case studies were conducted in NGOs that had undertaken an evaluation of a human service program. Key respondents with the required knowledge of the evaluation were identified through a research and evaluation manager in each organisation. Participants were interviewed about the context of the organisation, the program, how the evaluation had been conducted, how it had been disseminated, and what the impacts had been. Complimented by organisational documents, this information was analysed and presented as two case studies. Chapter five follows with the analysis from the presentation of these case studies, which is analysed using
Mark and Henry’s (2004) *evaluation influence* framework which were then subjected to further analysis to draw out the factors associated with practitioner learning from evaluation.
5. CHAPTER 5: FINDINGS FROM THE BRIGHTER FUTURES AND ST MARYS EVALUATIONS

5.1 Introduction

Drawing on organisational evaluation documents, as well as the accounts of human service practitioners and managers in the organisations running the programs, this chapter will summarise two evaluations of human service programs delivered by NGOs. The chapter will identify how the Mark and Henry (2004) framework is apparent in these accounts, which will enable the investigation of the factors related to practitioner leaning from evaluation. While all examples of evaluation influence are included, there is a special interest in those that resulted in or connected to practitioner learning.

Each of the evaluations is presented in three parts: the first part provides the context and key events of the evaluation drawn from organisational documents; detailing the organisation, the program, and the evaluation in order to provide a context for the following parts. The second part draws on Mark and Henry’s (2004) evaluation influence framework to present how the evaluation has influenced the program, drawing on both participants’ accounts of the influence of the evaluation and the organisational documents. This phase of the research enabled the exploration of practitioner learning, by identifying mechanisms that connect to learning in the context of these evaluations. The final part is the participants’ perspectives of the evaluation, which are illustrative of the factors important in fostering practitioner learning from evaluation. Evaluation influence captures the broad and diffuse effects an evaluation can have; taking this broader scope allows for the inclusion of mechanisms that may indirectly link to practitioner learning. To avoid interrupting
the narrative of cases, a more detailed summary of the data analysis process is included at Appendix E. The presentation of these factors as described by the participants will lead into chapter 6, which will present the conditions and barriers to practitioner learning from evaluation, as developed from the participants’ experiences of the evaluation discussed in the cases.

5.2 The Evaluation of Brighter Futures Delivered by the Benevolent Society

As described above, each case is presented in three parts. First, drawing on organisational documents, a case study of Brighter Futures delivered by The Benevolent Society will be presented (Section 5.2.1); summarising the organisation, the program, and how the evaluation was carried out. The next part (Section 5.2.2) will draw on the organisational documents and the participants’ accounts of the evaluation in order to present how evaluation influence (Mark & Henry, 2004) has played out. The final section (Section 5.2.3) will present the factors that seem to be important to evaluation influence playing out the way it did, through the presentation of participant experiences and perceptions of the evaluation. These perspectives are illustrative of the barriers and conditions to practitioner learning from evaluation that - combined with the findings from St Marys evaluation - will be expanded on in chapter 6.

5.2.1 Case study of the Brighter Futures evaluation

The Benevolent Society is a large NGO, with a highly professionalised workforce. The organisation provides services at 63 locations across Australia (primarily in
NSW), with 898 employees and 800 volunteers (The Benevolent Society, 2011).

Unusual amongst the other major NGOs, The Benevolent Society is non-religious, and has the distinction of being Australia's first charity. The central mission statement of the organisation is to "... educate, support and advocate for personal and societal change, to create a fair society where everybody thrives" (The Benevolent Society, 2012, p. 2). As such, the work of the organisation is centred on addressing barriers to full participation for disadvantaged groups; particularly through programs and services directed at children and families.

While always a large NGO, The Benevolent Society had been primarily Sydney based up until successfully tendering for the Brighter Futures program in 2006. The Benevolent Society is one of 14 NGOs in a mixed service model that also involves caseworkers from the NSW government department Community Services. The NGOs were referred to as ‘lead agencies’ in the evaluation documents. NGOs were asked to competitively tender for program contracts in areas across NSW, with The Benevolent Society successful in acquiring contracts for 51 local government areas, broken into five planning areas (Table 5.1). The Brighter Futures program was identified as being of interest to The Benevolent Society due to the evidence base behind it, its consistency with the organisation’s strategic focus on working with disadvantaged groups and areas and the early years of life.
Table 5.1 Local Government Areas where The Benevolent Society Delivers Brighter Futures (The Benevolent Society, 2012)

<table>
<thead>
<tr>
<th>Metro Central (North Sydney, Central Sydney, South-East Sydney)</th>
<th>Hornsby, Hunters Hill, Ku-ring-gai, Lane Cove, Manly, Mosman, North Sydney, Pittwater, Ryde, Warringah, Willoughby</th>
</tr>
</thead>
<tbody>
<tr>
<td>Metro South-West</td>
<td>Bankstown, Fairfield, Liverpool</td>
</tr>
<tr>
<td>Northern NSW</td>
<td>Armidale/Dumaresq, Gunnedah, Glen Innes Severn, Guyra, Gwydir, Inverell, Liverpool Plains, Moree Plains, Narrabri, Tamworth, Tenterfield, Uralla, Walcha</td>
</tr>
<tr>
<td>Hunter/Central Coast</td>
<td>Cessnock, Dundod, Maitland, Muswellbrook, Port Stephens, Singleton, Upper Hunter</td>
</tr>
<tr>
<td>Western NSW</td>
<td>Bathurst, Bland, Blayney, Cabonne, Cobar, Cowra, Forbes, Lachlan, Lithgow, Orange, Oberon, Mid-West Regional, Parkes, Weddin</td>
</tr>
</tbody>
</table>

The program expanded The Benevolent Society’s presence across NSW, with an increase in staff and administrative infrastructure. Expanding out to regional areas required some attention to the organisational position on indigenous issues, which led to the adoption of a reconciliation action plan (Benevolent Society, 2012). For regional areas in particular, Brighter Futures has enabled the organisation to deliver a number of additional programs at sites using the staff and resources in place.

Brighter Futures is an early intervention program funded by Community Services NSW targeted at families with children eight years or younger (or expecting a child) facing some of the following problems:

- domestic violence;
- parental drug and alcohol misuse;
- parental mental health issues;
- lack of extended family or social support;
- parents with significant learning and/or intellectual disabilities;
- child behaviour management problems; and
- lack of parenting skills (Community Services NSW, 2007).
Families voluntarily enter the program through referral by an NGO, or from the Community Services hotline. The service involved ongoing case-management with at least two other services (such as home visiting, parenting programs, quality child care) provided over the course of 18 to 24 months.

In terms of what constituted the program, Brighter Futures was the combination of service elements (i.e. case management, child care, home visiting, parenting program), and the eligibility criteria for clients. It was designed as an early intervention program, drawing on research evidence that suggests combining these elements can be an effective response to children at low to medium level of risk of harm from abuse and neglect. While the eligibility criteria had remained much the same, the referral process changed over the course of the program. The original guidelines worked on a quota of 80% referrals coming from Community Services, which changed over time to 60%, and eventually to no quota. These changes occurred in response to ongoing issues with getting an adequate number of referrals from Community Services, which restricted the number of community referral clients the program was able to admit (Benevolent Society, 2008).

The program had a number of desired outcomes around child development, namely bringing children in disadvantaged and difficult circumstances up to the developmental milestones of their peers. The logic of the program was built around the idea of using child focused services to help children catch up developmentally, while also working with families to improve parental capacity for supporting their children’s continuing development. From early on in the program it was well understood within The Benevolent Society that the families being referred into the program were on a very broad spectrum of need; families who really did not need the intensity and duration of service provided by Brighter Futures, through to families
that were on the high end of the spectrum, near the point of child protection
intervention who required much more intense work.

The program had a degree of adaptability to reflect the different contexts it
was delivered in, particularly regional differences and partnerships between the
NGOs and other agencies (e.g. local area health services). There was a lot of
flexibility in the implementation of the program; the evaluation report acknowledges
high levels of variability in the program across sites (Document D, p. xii). Some of
the differences were attributed to the partnerships of the sites, but also the different
relationships with local Community Services branches, and the types of staff
recruited. This in turn led to different referral and assessment processes, intervention
styles, and staff capacities. This was particularly so in the regional sites, where staff
were managed and structured by place rather than by program. The challenge of
geography was the main point of difference between the regional and metro sites,
having to work across many distant sites and at great distance from the local
Community Services office.

The program ran originally with the caseload split between the lead agencies
(NGOs) and Community Services, which was described as a mixed service model. In
2012 the decision was made to have the program delivered entirely by the lead
agencies. The decision about who provided the service was contentious with political
pressure for the service to be provided by NGOs, and union pressure to retain
government sector jobs. Up until this point, local Community Services centres
assessed and processed all referrals. This meant that the working relationship with
Community Services varied from area to area.

Across the program, interactions with local Community Services branches
centred on the referral of clients to the NGO to fill the required quota of helpline
referrals. Early on in the program there was collaboration between sites and Community Services on referrals, but over time this shifted, with appropriate referrals being determined by Community Services, limiting the input of program staff on the clients being referred to them. Due to staffing issues there were inadequate and often inappropriate referrals from Community Services.

5.2.1.1 The evaluation of Brighter Futures

The evaluation of Brighter Futures was an external evaluation contracted out to the Social Policy Research Centre (SPRC) at the University of New South Wales by Community Services. SPRC was contracted to undertake an evaluation of Brighter Futures across the many NGOs that delivered the program across NSW. The evaluation drew on three approaches; a process evaluation (which examined the conditions of service delivery), an outcome evaluation (which examined if the program achieved its stated aims), and an economic evaluation (an assessment of the economic impact of the program). The overall data collection included: service data collected across all Brighter Futures sites (client demographics, key vulnerabilities and service output data); the family survey (an instrument designed to measure family functioning, parental wellbeing, parenting skills, and child social/emotional development); a series of case studies; interviews with program staff and clients; and intensive study groups (participants involved in more extensive testing including developmental measures of the children).

An early part of The Benevolent Society research and evaluation team’s involvement was to examine the evidence base underlying the program to ensure the framework was compatible with The Benevolent Society’s organisational values. The Benevolent Society research and evaluation team was involved in some of the early meetings about the minimum data set (the demographic and output data The
Benevolent Society were required to provide to Community Services), the evaluation, and the form of the family survey.

The process evaluation acknowledged that “the aim of a process evaluation is to assist service providers to identify areas for change that can enhance service delivery” (Document K). The Benevolent Society sites were not included in the process evaluation; it included sites from other lead agencies delivering the program. There had been some initial discussions about the Bankstown site being included as a case study, but this did not eventuate. The outcomes evaluation was the part that included The Benevolent Society, and so is the focus of this study. The main data collection tool for the outcomes evaluation was the family survey, which was an extensive 32 page booklet of survey instruments. The survey was intended to be completed by families with a caseworker, but allowed for families to complete the survey by themselves in their own time. There was a concerted effort to try and ensure the person working through the survey with the family was not the regular caseworker for that family. The survey was conducted with clients two months into their involvement in the program, six months after that, and at the point of program exit. The family survey was designed to mirror the Family Strengths and Needs Assessment used by Community Services to measure client risks and strengths. It contains items on the following scales drawn from many existing clinical instruments (Document D§, pp. 258-259):

- parenting skills;
- hostility;
- parenting daily hassles;
- social network/support;
- mental health;
- self-esteem;

§ See Table 4.3 as a reference to the organisational documents cited in this case.
• drug and alcohol use;
• family cohesion/functioning;
• relationship satisfaction; and
• safe and nurturing environment.

The final report states that 2,563 families completed the survey, out of the 6,461 families in the Brighter Futures system. The attrition rate for participation in the evaluation was extremely high. The final report suggests that only 30% of respondents completed more than one survey, and that the third follow up only had 150 respondents (Document D).

Along with the family survey, service data from the operations of Brighter Futures was used in the evaluation. From the start of the program, the data set was sent back to Community Services, the collection was highly prescriptive and detailed, and the requirement to collect this data was written into the service contract. This data included demographics and service delivery data:

• how many families used the service;
• how many children in childcare;
• how many parents in programs;
• how many children in playgroups;
• the main issues (e.g. domestic violence, mental illness);
• service costs per family;
• start to exit date; and
• reasons for leaving the service.

The collection of this data was initially paper based across all the lead agencies, making it difficult and time consuming to administer. Once it was understood that Community Services would not be sharing this data with the lead agencies, it had to be collected twice in order for The Benevolent Society to obtain any operational information.
The key findings of the outcome evaluation in the final report (Document D) were that the program seemed to be meeting client needs and modestly improving outcomes for families. There was recognition that many families did not benefit from the program, and that some of the reductions in risk also occurred for the comparison families (from the small intensive study group) who received no intervention. The evaluation reported a relationship between families' duration in the program and the achievement of case plan goals, while also acknowledging that the majority of families did not remain in the program for the entire two years.

The process evaluation highlighted the challenges involved in the implementation of such a large program, and some inefficiency in the program model that required attention. The evaluators acknowledged the level of variability in service delivery, highlighting the difference in service models used by the lead agencies and Community Services. Difficulties in the partnerships between the lead agencies and local Community Services branches were also identified, specifically that there was an underlying inequity in the arrangement that limited NGO involvement in decision-making. There was also some recognition that a large proportion of the families entered the program with acute issues outside the child/parent development thrust of the program. The evaluation explicitly acknowledged that the program involved a client group with issues beyond what the program had been developed to work with.

The final evaluation report recommended changes to entry pathways to improve efficiency and client waiting times, acknowledging the findings of the Keep Them Safe report (Department of Premier and Cabinet, 2009). Increased flexibility to the service model was also recommended. Some families did not need case management, while other families with more complex needs required timely referral
to appropriate services. It was also suggested that other agencies could be brought into the partnership. The evaluation report acknowledged the limitations in the data collected, namely: the short timeframe, the unreliability of some data, the lack of a comparison group for the family survey, and the unrepresentativeness of the sample.

Dissemination of the evaluation findings to practitioners relied heavily on the Benevolent Society research and evaluation team extracting important issues from the data for practice, and promoting discussion about change among the management group. The Benevolent Society had initially been waiting for access to lead agency and site specific data to use for program improvement. After being told that there would not be any site level analysis or feeding back of specific results to lead agencies, The Benevolent Society began discussions with SPRC to produce some site specific analysis. This became a complicated three way negotiation between all of the parties. Community Services was very clear that site level analysis was not going to be done and that The Benevolent Society would have to talk to SPRC about it, while SPRC did not wish to do any additional analysis that had not been done for Community Services. Despite The Benevolent Society practitioners collecting the surveys, there was a strong sense of ownership of the data within Community Services, and a general objection to The Benevolent Society getting access to this data. Despite this, the process began with contracting SPRC to produce reports for The Benevolent Society, which took about a year of negotiation. Community Services signed off on an internal report and a research snapshot that was published late in 2012.
5.2.2 Evaluation influence in the Brighter Futures evaluation

Mark and Henry’s (2004) *evaluation influence* framework was applied to the Brighter Futures evaluation, drawing on the information from the organisational documents discussed above, and the interviews with participants. All identified impacts of the evaluation were included, and then analysed through the prism provided by the framework, which included mechanisms not generally associated with practitioner learning.

For a number of events included it was not completely clear that the evaluation was the source of the influence, for example some of the high level policy changes that had a number of influences beyond the evaluation. All examples were included if a participant or a document indicated the evaluation may have had an influence, even indirectly.

The examination of influence in the case was also complicated by the involvement of external agencies, namely Community Services as the funder and co-deliverer of the service, and the SPRC as the evaluators. While there was a lot of detail available on some of the internal changes, for the big picture changes that affected the delivery of the program it was often unclear what role the evaluation had played in influencing change. Particularly in terms of the change to the program being delivered purely by the lead agencies, it was unclear if the evaluation had influenced this change, or if this was a decision already made that had drawn on the evaluation for support. A separation had been made between internal policy
change/learning and external change, reflecting the fact that the influence of the evaluation played out beyond the organisation being studied.

Table 5.2 presents the influence of the Brighter Futures evaluation mapped out onto the Mark and Henry (2004) evaluation influence framework. Not all of the mechanisms in the framework were identified in this case (e.g. elaboration).

**Table 5.2 Mechanisms Reported in the Brighter Futures Evaluation**

<table>
<thead>
<tr>
<th>Level of Analysis</th>
<th>Mechanisms</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Individual</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>General</td>
<td>General Heuristics</td>
<td>• The evaluation confirmed practitioners’ expectations that the program was working;</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• The evaluation fostered a sense that the program had a future because it had been evaluated.</td>
</tr>
<tr>
<td></td>
<td>Priming</td>
<td>• Practitioners talked about getting a sense of how their role connected to the rest of the program, a sense of the logic of the program as a whole.</td>
</tr>
<tr>
<td></td>
<td>Salience</td>
<td>• Some of the evaluation information highlighted characteristics of clients that practitioners reported being more responsive to and proactive in identifying.</td>
</tr>
<tr>
<td></td>
<td>Opinion /Attitude Valence</td>
<td>• Changed attitudes about evaluation being negative, that it can be done in a way that supports practice and reflection;</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Engendered resentment towards Community Services as practitioners did not think the comparison part of the evaluation properly represented the difference between their work.</td>
</tr>
<tr>
<td><strong>Interpersonal</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>General</td>
<td>Change Agent</td>
<td>• One of the managers actively engaged in attempting to foster a culture of evaluation among practitioners, to have them be excited and critical about evaluation.</td>
</tr>
<tr>
<td>Cognitive / Affective</td>
<td>Local Descriptive Norms</td>
<td>• A culture of evaluation and change practice developed around the organisation affecting attitudes about the role of evaluation in practice.</td>
</tr>
<tr>
<td></td>
<td>Injunctive Norms</td>
<td>• The evaluation affected the delivery of the program to be more standardised, meaning that if particular checklist requirements were not fulfilled a service was not counted as delivered;</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Managers were influenced to be aware of and informed by the data in the way they ran their sites, to understand the implications and be responsive to issues;</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Practitioners were encouraged to think about and use data in their practice through managers, supervision, and participation in the Program Improvement Plan process.</td>
</tr>
</tbody>
</table>
### 5.2.2.1 Individual level change

Some of the individual level change seemed to reflect *heuristics*, which Fleming (2011) describes as a mode of attitude change that bypasses systematic thinking in favour of following heuristic rules (e.g. long messages are valid messages). While the term heuristic may seem derogatory, it simply reflected in this case a consistency...
between the evaluation information and the internal values of a person. Participants described the main practice level impact as confirming their expectations that the program was working, which served to encourage and validate the work of staff:

...we know anecdotally and through our own information that it’s having an impact, so to have something official by academics that’s been rigorously assessed... that’s a terrific thing, and that the program will continue and possibly be extended is only good (Courtney, BF Managers Group).

This was connected with an increased sense that the program was likely to continue and be sustained, and that their employment and the program were more secure because it had been evaluated.

*Priming* describes the effect an evaluation can have in highlighting a concept and making it central to the judgement of an issue (Oliver, 2008). Some participants reported a greater sense of a program level view, an improved understanding of other sites and contexts, what the big picture outcomes were, and how their work fitted into the whole of the program “...sometimes the temptation is just to think that you’re working with ten families and that’s all that’s going on in the world and you can’t sometimes... see the broader impact” (Courtney, BF Managers Group). One participant suggested this program level view was also thought to lead to a greater sense of model fidelity:

...I do think it’s created a more thoughtful approach to our work, and I suppose a sense of model fidelity in terms of providing how the four different sections interrelated with one another to create an outcome for a client (Megan, BF Managers Group).

Participants suggested the evaluation helped staff see past the barriers and challenges in their day-to-day work and to understand these in the context of the whole program.

*Salience* is the increased importance of an issue that affects the way a program is understood (Henry & Mark, 2003). Participants expressed surprise at how many families had been through the program, and how common some of the key vulnerabilities were. It was observed that simple information like demographics
altered practice, it could make staff more responsive to social cues, and particularly how common issues like drug abuse and domestic violence were among the families:

...she’s here and I know there’s a 50 percent chance there’s domestic violence [referring to demographic information about the clients collected for the evaluation], so I just need to keep my eyes and ears open and just give an opportunity or ask the point blank question “is this happening for you and how do we talk about that” (Megan, BF Managers Group).

Respondents described how the evaluation and the workshops reporting back on the findings led to some deep discussions around theories of change and the different assumptions implicit in disciplines (e.g. psychology, social work, community development).

Opinion/attitude valence refers to the change of beliefs and attitudes about a program through a thoughtful process of reflection and learning (Fleming, 2011; Henry & Mark, 2003; Oliver, 2008). While the evaluation did not really challenge what individuals thought about the program, being subject to a large scale evaluation influenced practitioner attitudes and capacity to support evaluation. It was seen as significant that evaluation could occur and that people did not necessarily get ‘beaten up’ by the results “that’s important, you can have evaluation and have it mean that you are curious about your work and that its okay to have some difficulties in your work and for that to be on display if you like” (Vanessa, BF Managers Group). The increased organisational capacity and support for evaluation meant that some practitioners were engaged in the discussion of evaluation findings and problem solving around issues identified by the evaluation. Participation in these processes seemed to change attitudes about the relevance of evaluation to practice.

Conversely many participants expressed concerns about the integrity of the evaluation because of the quality of the section which compared lead agency service delivery to Community Services delivery of services. This was described as
stemming from the late inclusion of this part of the evaluation, and the findings that suggested that Community Services practitioners were more qualified and handled the more challenging end of the Brighter Futures caseload. Practitioners expressed some frustration with the way their work was represented in this part of the evaluation and felt that the comparison was politically motivated. This fostered some resentment against Community Services and the potential conflict of interest they had in evaluating the program while also delivering it:

...so there have been some tense meetings about these, because there is conflicting interests in Community Services as they’re funding us to do it, but they’re providing it as well and that’s been a challenge for them, and then there’s pressure from government for it all to go to, sorry from politicians for it all to go to NGOs, but then there’s also the PSU pressure, the Public Service Union, pressure to keep it within Community Services (Megan, BF Managers Group).

5.2.2.2  Interpersonal level change

One of the participants identified herself as a change agent, someone who as a result of the evaluation focused on changing how things were done in the organisation (Mark & Henry, 2003). This person was proactive about building a culture of evaluation in her site, working to engage staff by promoting an open dialogue about what the evaluation was for and encouraged practitioners to contribute through sharing any difficulties with the process they had:

I was trying to be a person who said, so if we had some questions or worries can we put them on the table and can you answer them, so that the staff could see we were having an open dialogue, we could get an open response and feel okay about going forward (Megan, BF Managers Group).

Participants felt that while the evaluation did not really change what people thought about the program, it did play a role in developing a culture of evaluation and change practice around the organisation. This seems to be akin to the mechanism local descriptive norms, which Oliver (2008) describes as evaluation affecting the

---

6 This union is now called the Public Sector Association.
social behaviour and interaction of stakeholders. Beyond just the organisational systems in place, the evaluation developed capacity and engaged managers and practitioners in the discussion of data, leading to attempts to solve practice level problems.

Through *injunctive social norms*, which Oliver (2008) describes as agreed upon principles about how to conduct oneself in a particular setting, the evaluation had some influence in the delivery of the program. Largely interpreted as an example of the evaluation promoting a reductive view of practice, practitioners were critical of practice having to be done in a more standardised way in order to be counted for the evaluation “*the rules and regulations they’ve wanted to put around some of those that has impacted on the delivery...you have to do it this way for the evaluation*” (Megan, BF Managers Group). Injunctive norms were also changed in encouraging both Brighter Futures managers and practitioners to reflect on evaluation information. For managers that meant an expectation that they were aware of and used evaluation data in the management of their site:

*...there’s an expectation that you’ll not only read it, but know the implications for your service, you’ll have a program improvement plan, you’ll have your own program improvement plan to help increase your retention rates... so a sense that you would need to respond to what the data the evaluation data and indeed throughput data was telling you* (Vanessa, BF Managers Group).

There was also a reported change in the injunctive norms for practitioners, that support and attempts to engage them in evaluation came with an expectation that practitioners think about and be aware of the implications of evaluation findings for practice:

*...sometimes it (evaluation) feels like it’s just existing out there, and what I liked about it was it brought it back to ‘hey this is what you’re doing... this is the things of what we’re seeing’ I guess it just made us think more about our practice and tried to encourage us to be more purposeful in what we’re doing...* (Natalie, Team Leader).
Some participants thought that simply raising and discussing the issue among practitioners often led to measurable improvements; for example, after a report back on the findings which highlighted the issue of clients disengaging prematurely staff were noted to be making an extra effort to work with families who appeared to be at risk of disengaging. The ability to use evaluation to resolve practice issues was seen as positive:

...for example, and one manager said, we had nine exits in the last quarter, and when we started focusing on really persistently if people were really looking like they were going to exit in the first three months, we started really focusing on them and helping them stay connected, we only had two exists in the last quarter, and she was really proud that she was able to say that, and it was like yeah good on you (Megan, BF Managers Group).

5.2.2.3 Collective level change

Much of the change that resulted from the Brighter Futures evaluation impacted at the collective level in changes to policy and organisational processes. Some of the participants indicated that the evaluation had highlighted some key operational issues which led to policy consideration at The Benevolent Society level. The organisation also had used the evaluation data in tendering for other contracts.

Early exits and client engagement were identified as particular challenges for the program, particularly in regional sites and among helpline referrals; families tended to get involved in the program during a crisis, and then leave once their short term issue was resolved “we were having families coming in and once their crisis had been resolved they were quick to leave the program... whereas we had intended families to stay with us for about two years...” (Lauren, BF Managers Group). A number of processes (described below) were put in place to address this.

The evaluation led to agenda setting in the Brighter Futures managers group. Participants noted that the evaluation put data behind some of the perceived differences between sites, which resulted in more productive discussions at the
Brighter Futures management level. Managers were expected to not only understand the implications of data for their own site, but to contribute to discussions about other sites “...they get their data and have to account for their data if you like, they would all know each other’s data as well, if you were managing your program in Bankstown/Liverpool/Fairfield, you’d probably know how North Sydney was going...” (Vanessa, BF Managers Group). Through the Brighter Futures meetings, managers were in the position of having to constantly think about and use the data, and report back on what changes they had made and what the effects had been. There was a standing agenda item at these meetings about the evaluation, with managers expected to talk through the implications of the data for their service sites:

Many of the changes that occurred had their agency outside of The Benevolent Society, with some decisions made within Community Services. As some examples of influence seemed to be located outside of The Benevolent Society, for the purpose of this analysis some of the mechanisms of the Mark and Henry (2004) framework have been separated into internal and external. Within The Benevolent Society there were some examples of policy-oriented learning (internal), which resulted from the evaluation. The issue of early exits led to some critical discussion about whether families needed to be in the program for two years to get the benefits, whether there could have been more flexibility around exiting:

...we’re looking at how we capture around our Aboriginal families, and getting it identified that families don’t necessarily need to be with the program for two years in order to meet the outcomes they’re looking for and finish off their family plans (Stephanie, Team Leader).

The focus on early exists also led to a high level of interest in why families were exiting, in order to identify at the point of assessment families that are likely to engage well.
An issue that was put on the agenda because of the evaluation which lead to internal policy oriented learning was the effect the client base was having on the practitioners. Staff had been hired to deliver an early intervention program, but were exposed to families much higher on the spectrum of risk. At one level this was problematic because the program no longer matched the evidence base that had been developed for it, but it also brought up a number of other issues around staffing:

...we recruited staff at this level thinking that this program was going to be an early intervention program... whereas in fact what we were seeing was families very close to the tip of child protection, very complex issues surrounding them, and I believe we had staff in our region that were finding it very very difficult to understand how you actually manage that (Lauren, BF Managers Group).

This seemed to be linked with issues of staff burnout and turnover:

I think we have been exposed to around eleven child deaths since we started doing Brighter Futures and they weren’t homicides at all. They were you know fire or those kinds of things and that just reflects the social gradient, that child deaths occur and accidental deaths occur in these families much more frequently because of the lifestyle issues and the sheer impoverishment the families live in, and they take their toll on the staff as well (Vanessa, BF Managers Group).

This led to some reflection about The Benevolent Society’s responsibilities as an employer, and how to equip staff with the resilience and judgement needed to work in this environment. The evaluation acknowledged that the client group did not match what the program was set up for, that there was a need to improve the targeting of families that were likely to benefit, and a need for flexible service delivery depending on the level of need in the client family.

Participants were surprised to find that the evaluation was using the achievement of case plan goals as data, as these were developed quite informally in discussion with the client. Once it was understood that case plan goals were being used in this way, it began a robust discussion within the program about what a good case plan goal looked like, and the sharing of research on how to write them.
Another issue the evaluation brought into clear focus for The Benevolent Society was the diversity in the delivery of the program. While managers were engaged in attempting to understand the different program contexts through the managers’ meetings, recognition of the very different contexts of regional and metropolitan service delivery resulted in internal policy oriented learning. While many of the participants felt that Brighter Futures as a program was quite flexible in its implementation, there was still a sense that more had to be done in addressing the challenges of regional service delivery:

... we’re still thinking we can do home visits in the way we assume home visits should work, so by and large we’re still trying to do those things, have we compensated enough for say rural areas, where it’s not that feasible unless you live in the town to use childcare as a mechanism, have we really worked out how what to do about that, I’m not sure about that” (Vanessa, BF Managers Group).

The evaluation also contributed to external policy-oriented learning through the use of evaluation data in The Benevolent Society’s submission in reply to the Wood report (The Benevolent Society, 2008). The data highlighted problems with the Community Services side of Brighter Futures, specifically the inability to provide the NGOs with the required caseload referrals. The impact of this was fairly significant with recommendations for the increased funding and delivery of services by NGOs. In their response to the report Community Services indicated the organisation would wait for the completion of the Brighter Futures evaluation before making decisions about changes to the Brighter Futures model (Department of Premier and Cabinet, 2009). The impact of this on the decision to change the service to being NGO run was unclear from the perspective of the participants, therefore the use of the evaluation data in the submission had been treated as policy-oriented learning rather than program change.
There was some examples of *program continuation, cessation, or change*, which in the framework is considered distinct from policy change (Oliver, 2008). The evaluation identified the limited evidence base and efficacy of some of the parenting programs being delivered, as a result these were changed “...*I do think the parenting program issue that came out in the evaluation, that group parenting programs weren’t particularly successful, and I think they’ve disappeared virtually straight away*” (Vanessa, BF Managers Group). As a result of some of the evaluation recommendations the service provision guidelines changed engagement times from two years to one year. This recognised the need for program flexibility to reflect the different types of families served; and that the benefits of the program could be realised by some of the families without needing the extra time. These program level changes occurred alongside a number of significant internal and external policy changes.

Within The Benevolent Society a number of *policy changes (internal)* occurred as a result of the evaluation. The evaluation led to an increase of the organisation’s evaluation capacity through the hiring of evaluation staff and the creation of a Research to Practice Manager position. This increased capacity within the team leading to some changes in the way information was disseminated. Previously the research and evaluation team in The Benevolent Society would present to the Brighter Futures managers who were in turn expected to present to staff, but often these managers did not feel well placed to feed the information back to their practitioners:

*The team had a lot of questions and I was like I’m not sure, I didn’t feel as though I had all the information to give them... I don’t think that myself or my manager did a great job in actually feeding information back to the guys that are actually doing the work. That’s something that in hindsight we could have done more actively, whether that’s us ourselves having more time with*
Participants also remarked that there was a lack of continuity in the management group that may have limited the feedback to practitioners. From surveys of the organisation’s evaluation capacity, it was found that staff felt that evaluation recommendations were not being implemented. This was attributed to conflicting perspectives on who should be driving improvement processes in the organisation; it was recognised that leaving the information for the operational side of the organisation was not working:

*We weren’t getting very positive feedback that evaluation recommendations were going anywhere or improving practice, and we had a bit of a debate with my boss about what role if any should we have in that, and I didn’t really think that it was a role for us, but she was of the opinion that if we don’t do it then nobody will do it... just leaving it to the operational side of the organisation... was obviously not working* (Victor, BF Managers Group).

This then led into the development of facilitated workshops with managers and practitioners to help develop responses to evaluation findings linked to changes in practice.

Over the course of the evaluation a number of guidelines were developed to guide The Benevolent Society's response to evaluation information. The research and evaluation team developed an evaluation policy that sets out how evaluations should be undertaken and disseminated in the organisation (Document L). The experience of the Brighter Futures evaluation was highly formative in the creation of the policy; it emphasised ethical, useful, participatory, and credible evaluation. Another significant development was the use of Performance Improvement Plans (PIP) as a way to get practitioners to reflect on evaluation findings and develop change strategies grounded in practice. The process involved a workshop with staff in order to complete the template, identifying the implications of the findings for practice, areas of change, and indicators to measure the effects of change. PIPs were monitored by
the research and evaluation team with reviews at six and twelve months, followed by a report to the Senior Executive Team. Workshops and the PIP process were used to facilitate collaborative problem solving, and share practices around the different sites.

The extent to which the evaluation influenced *external policy changes* was much more speculative on the part of the participants compared to internal policy changes. Some participants thought the evaluation had been highly influential at the Community Services level, while others thought that the evaluation was used to support existing decisions made about the program “*It seemed to me that they were trying to argue a case and look for any support for that case rather than just say what the data said*” (Earl, BF Managers Group). Three intertwined external policy changes occurred that had a significant impact for The Benevolent Society’s delivery of Brighter Futures: over time the required percentage of Community Service referred cases dealt with by The Benevolent Society decreased; the Community Services side of Brighter Futures took greater responsibility for serving the high end intensive cases; and Brighter Futures being delivered solely by the lead agencies. Initially, eighty percent of referrals to Brighter Futures were to come through Community Services. This percentage reduced over time as Community Services were often unable to provide the number of referrals required:

*...that was something that was found in the evaluation, that we were always receiving more community referrals than helpline, even though we were targeted to do it the other way around... that’s part of the new guidelines, there is no distinction* (Monica, Team Leader).

The evaluation highlighted the complexity of some of the cases dealt with, which led to Community Services taking on the high end caseload of families that were beyond the specifications of an early intervention program “*Community Services have still retained, they now call it Strengthening Families, which is their*
kind of bit of Brighter Futures, but they’re getting the families (on) the child protection end” (Victor, BF Managers Group).

The evaluation also played a role in the decision to hand the delivery of the program over to the lead agencies. The transition to a solely lead agency delivered service meant an increase in placement numbers, and an overall decrease in the number of cases dealt with by Community Services in the program. Part of this handover meant that The Benevolent Society became responsible for their own referrals and assessments, previously assessments were done by the local Community Services branch “...the person at the (Community Services site) had been the person to give the okay or not on community referrals, so whenever we received a community referral we’ve had to send it back through that manager” (Monica, Team Leader).

The next section explores the participants’ experiences of and perceptions of the evaluation in order to explore why the evaluation was influential in this case. These accounts of the evaluation are descriptive of the conditions and barriers to practitioner learning; describing how the evaluation could have been more influential, and what things were important for the evaluation to result in change to human service practice.

5.2.3 Participant perceptions of the Brighter Futures evaluation

The participants’ perceptions and experiences of the evaluation were illustrative of the conditions and barriers for practitioner learning that will be outlined in chapter 6. These are presented in this section as a number of sub-categories: the context of
working with Community Services in delivering Brighter Futures; how evaluation connected with practice in The Benevolent Society and Brighter Futures; how participants experienced the evaluation; and the reflections participants had about how evaluation could promote practitioner learning.

5.2.3.1 **Context of Brighter Futures with Community Services**

This section provides a sense of the participants’ experiences of working in partnership with Community Services in delivering Brighter Futures. This foreshadows some of the challenges The Benevolent Society had in working with Community Services in the evaluation and provides some insight into the barriers and conditions that enable practitioner learning from evaluation. Participants emphasised the challenges of being in a program that was so connected and even reliant on Community Services “it has been difficult to influence practice in the community services sector, because of this entrenched headspace they (Community Services) have, it’s really difficult for them to... actually change practice” (Lauren, BF Managers Group). Some of the sites reported a fraught relationship with the branch of Community Services they were partnered with. There was sometimes a lack of clarity around the roles and responsibilities at the caseworker level, despite attempts to foster collaboration at the management level. One participant described working with Community Services as challenging due to the child protection mindset in the organisation, and the pernicious view that early intervention (i.e. Brighter Futures) was low level work, both for their own staff and for the lead agencies “well they’re not really working they’re just fluffing about with these families, the real work is child protection, so I don’t think from this region there’s been a great commitment (from Community Services)” (Lauren, BF Managers Group). While there was a common perception that Community Services dealt with the more challenging families because they had a larger load of helpline referrals, participants felt that
many of the community referrals, upon full assessment, presented with the same issues and levels of risk.

Across some of the sites The Benevolent Society provided some of the Brighter Futures services to Community Services clients (i.e. childcare). In these sites the relationship was mostly positive, but there was sometimes a sense from staff, that Community Services staff felt they were owed a level of service by the partner organisation “... I find some arrogance at times from the (Community Services) staff... it was a sense of entitlement that we’re DOCS, we want five days a week of childcare and that’s what you’re going to deliver for us” (Stephanie, Team Leader). The Benevolent Society staff had reported some success in dealing with this, in making the conditions of service clear:

...it’s been a matter of working with each of the teams and one on one at times, and sort of building a relationship with them [Community Services] so they know, we can do whatever it is you need us to do but we’re all in this together and no one is better than anyone else (Stephanie, Team Leader).

Some participants reported a positive working relationship with Community Services, particularly valuing the ability to have an off the record discussion about case matters like the appropriateness of referrals. Consistency in staff at Community Services was suggested as an important pre-cursor for an effective working relationship, as good relations were something that developed over time "I think we have a good relationship with Community Services and that's because we worked at that... we're also co-located with them... made sure we've had regular meetings with them, and made sure there's good communications" (Christine, Family Worker). The evaluation identified issues with the partnerships between Community Services and the lead agencies, suggesting that where they worked well it tended to be due to the initiative and personality of individuals rather than the program framework.

Community Services (New South Wales) is often known as DoCS (Department of Community Services).
While both Community Services and The Benevolent Society had difficulty attracting and retaining qualified workers, respondents reported the Community Services staff performed much of the same work as the NGO practitioners, although with better pay and conditions. This meant lead agency workers often moved over to Community Services as they gained experience. On the other hand participants were aware of the possibility of the program being handed over to the lead agencies, resulting in job insecurity among staff in Community Services.

It was observed that it was potentially risky for The Benevolent Society to be partnered with Community Services in delivering Brighter Futures, particularly so in areas with a large Aboriginal community:

...for us to be seen in a relationship with Community Services is not in our best interest in our communities, because we already are needing to present ourselves as a sole entity, a different entity than Community Services... within the Aboriginal community, for us to be seen in a partnership with Community Services, is not in our best interests... while they (Community Services) see that as really helpful, to break down some of the entrenched views that communities have of Community Services... but for us, it's hard to overcome people thinking that we are actually a part of Community Services and what that then leaves people with about removal of children and historical facts...

(Lauren, BF Managers Group).

5.2.3.2 Perceptions of the culture of evaluation in The Benevolent Society

The organisation had a strong stated commitment to evidence based practice, which was understood to include the evaluation of programs:

The Benevolent Society (TBS) ensures that its services are informed by the best available evidence of what works, along with the knowledge and experience of practitioners and the views and experiences of clients. Evaluation is an important source of this evidence (Document L).

When asked about how evidence based practice was reflected in The Benevolent Society, participants spoke about the use of evidence based assessments, choosing to pursuing service contracts for programs that have an evidence base, efforts internally to report back on evaluation findings regularly, and formal processes around the
discussion and reflection on evaluation findings as a way of developing
recommendations "I don't like to do anything personally in my practical work that
doesn't have an evidence base, so I rely heavily on Benevolent Society to direct me
with that... and promote only programs that have an evidence base..." (Christine,
Family Worker). Some sites allocated reading time for practitioners, allowing them
to keep abreast of the literature in their disciplines. Having the capacity within The
Benevolent Society to support evaluation and identify issues for practice was highly
valued:

*"I don't know too many other places that have that [in-house evaluation
capacity]), that in itself provides a recognition of the work that we do and
how we try to influence change, so it was about getting someone, the likes of
an [Research Manager] to digest all that information and put it in a form that
people could actually understand and take on board"* (Lauren, BF Managers
Group).

Participants provided examples of how evidence based practice was
implemented in the Brighter Futures program. Prior to the beginning of the outcomes
evaluation (i.e. the family survey) The Benevolent Society undertook a number of
evaluation activities as part of their own commitment to evidence based practice.
Early in the operation of Brighter Futures, participants reported a program logic
model (e.g. Cooksey, Gill, & Kelly, 2001) being undertaken in order to get a sense of
the job based requirements of staff, and how to support knowledge sharing and skill
development. The program logic also laid out how the program might be evaluated
internally, identifying the important outcomes and how these might be measured.
Despite a lot of effort going into it, participants reported the program logic had not
really been used, partially because it was understood the Community Services
evaluation was congruent with the types of outcomes and measures thought to be
important within The Benevolent Society “...it was (the evaluation) covering a lot of
the outcomes and areas we wanted to capture anyway” (Victor, BF Managers
Group). The organisation also expanded its internal research and evaluation capacity by taking on additional research staff, and the creation of a Research to Practice Senior Manager position.

However, there was a sense that the level of implementation varied across the organisation and that evidence based practice was not well resourced enough for it to occur in ongoing and sustainable way:

... we’ve always had a value or kind of strategy around evidence informed practice or evidence based practice, the extent to which that gets implemented varies considerably, and I think we probably haven’t put the resources in as an organisation enough to actually happen (Victor, BF Managers Team).

5.2.3.3 Experiences of the evaluation
The participant experiences of the evaluation revolved around perceived shortcomings with the evaluation. Participants detailed concerns about data quality and credibility. They felt the evaluation relied on data that did not represent their work well and lacked credibility. They questioned the relevance of the evaluation, that the evaluation lacked connection to the practice context and the way the program was experienced by practitioners. The evaluation was thought to be quite invasive in the way it interfered with practice and required extra work from practitioners with no perceived benefit. The management of the process was also criticised by practitioners, highlighting the lack of consultation and poor management on the part of Community Services. Finally, the lack of effective dissemination, resulting in limited knowledge of the evaluation at the practice level was noted.

- Data quality and credibility
The evaluation highlighted within the organisation the need for credible data to inform practice and policy. There was cynicism about the use of the Brighter Futures data to inform practice because of problems with the way it was collected “No, it just seemed to be an inadequate evaluation mechanism upfront, and then an inadequate
representative volume as it was rolled out” (Earl, BF Managers Group). Some of the participants thought the evaluation would be more stringent, involve more objective criteria, and have a control group. There was a lot of discussion by participants about problems with data quality, both in terms of the minimum data set and the family survey. The problems with the service level data were mainly associated with many staff entering it “I think their data was imprecise, so the definitions when you’re working across the state and across different personnel, lo and behold people enter, understand things differently and enter incorrect information” (Earl, BF Managers Group). Some of the participants reported that the evaluation data lacked quality because of the low rates of surveys completed in particular areas, and that overall the targets for the number of surveys completed were not reached.

The survey data itself was thought to be of poor quality by the participants due to the response bias associated with using program staff as defacto-researchers “...we had grave reservations about it, that a caseworker interviewing a client verbally for feedback about the service... there’s going to be a biased response” (Earl, BF Manager Group). The length and language of the survey was also thought to be problematic, particularly among communities with English as a second language. Participants were also critical of asking parents to complete measures about the development of their children, which was also thought to be open to bias. That a program focusing on improving children’s development outcomes was evaluated by using the parent’s perception of their child’s progress (along with their own parental capacity) was seen as a poor approach to evaluating the service. The evaluation did include an intensive study group where child developmental changes as a result of the program were studied. However this was a relatively small number of participants across the Brighter Futures evaluation as a whole. The outcome
Evaluation was seen to be limited in only presenting part of the picture of what the program was about:

...we’re assuming that if parents are parenting better then they’ll have a sense of their own self-efficacy as a parent which will mean that the children are doing better, but on a whole we’re not measuring the children’s developmental outcomes (Vanessa, BF Managers Group).

Overall, participants observed that there was a lack of a focus on children’s experience of the program, as well as on their outcomes.

The evaluation included a number of measures which participants were dubious about. For example the evaluation used a measure ‘case plan goal achieved’, which the practitioners initially did not know was being used. Case plan goals came about informally through discussion with clients, and they reflected a variety of subjective goals based on the expectations and life situations of clients. Capturing these for use as a common metric was thought to be problematic as they varied significantly, particularly across families at different levels of risk:

...we’re a bit like, so did you tell us you were using that in that evaluation, because we would have said okay what are good case plan goals, how were you using them, what do they look like... so that was for us a bit unfortunate (Megan, BF Managers Group).

Another criticism of data collection was that it treated families that left the service early as early exits, even when they left the program due to moving to a different area. It was suggested that the high rates of Aboriginal families leaving the program early could be partially explained by the high level of transience in those areas. These families would often connect to a Brighter Futures service in another area, which was not recognised in the data.

In terms of practitioners’ assessments of the evaluation and the likelihood of data influencing practice, participants highlighted that the background of practitioners can mean significantly different interpretations of the program and the role of evidence and evaluation. Disciplines draw on different types of evidence and
understand the role of evaluation in different ways, and this played out in the Brighter Futures sites:

...and I think the biggest challenge is them implementing it to a great variety of staff, so we've got staff that are social work trained, psych trained, social science trained, early childhood trained, how do you make sure there's a consistent absorption of that information, then application, I think that's the challenge (Megan, BF Managers Group).

At one site in particular there was tension between a set of professions (e.g. speech pathologists, psychologists, & occupational therapists) and the teaching staff in terms of understanding the program and the nature of evidence. The personal approaches of some practitioners like teachers and case workers clashed with the use of standardised assessment tools, along with the requirements of program fidelity for the evaluation "it can sometimes be challenging in convincing the teachers that its valuable to put a child through a standardised assessment... the teacher sees the children from a different perspective" (Monica, Team Leader). One participant thought the challenge in getting engagement with evaluation and evidence was less about discipline, but more about the stability and continuity of staff. It was suggested that much of the shared learning, understanding and values exchange was lost when there was high staff turnover.

Along with issues of data quality, the credibility of the evaluation was challenged by cynicism among practitioners about why the evaluation was being conducted, and the power Community Services had in determining the evaluation direction. Among managers it was understood that the purpose of the evaluation was to justify the investment of government funds and to secure ongoing funds from treasury for the program to continue and expand. Indeed several participants used this to convince staff to comply with the evaluation. Fostering program learning was not perceived by the staff to be on the agenda “Um no, I don’t think Community
Services cared much about the individual lead agency data in terms of them 
commissioning [the evaluators] to do an evaluation piece” (Vanessa, BF Managers Group). For this reason, some of the participants were quite surprised that I had identified the Brighter Futures evaluation as a case study to explore practice level learning.

Participants expressed cynicism about the purpose of one of the most controversial parts of the evaluation; the inclusion of a comparison between lead agency and Community Services delivery of the program. Participants took the view that the decision to add-on a service delivery comparison was driven by political considerations at the Community Services level. There was also a lot of cynicism about some of the findings and the influence Community Services had on the evaluation. There was a sense that this comparison was not in the spirit of the original evaluation, and yet it came to be added due to political imperatives. It was understood in The Benevolent Society that the evaluation would inform the decision-making about the split of services, or the possibility of the service becoming wholly lead agency delivered.

- Perceived relevance of the evaluation

Some of the concern about the evaluation had to do with the way it represented the implementation of the program, which participants felt differed from their experiences. It was recognised that the program was delivering services to a much broader group than the early intervention families described in the manual:

I think the old DG used to call it swamp draining, that the only way you’d ever get the number of eligible cases sitting inside Community Services ever attended to is if you started to meet the needs of this client group, or this client group, being the ones that were at the higher end of intervention needs (Vanessa, BF Managers Group).
Partly this was thought to be addressed by the higher end or more complex cases being taken on by Community Services. However many of the participants felt that this was not so, and that inappropriate referrals often came through to The Benevolent Society and the other lead agencies. One participant thought 'swamp-draining' was indicative of a fragmented and piecemeal community services sector; that there needed to be an interagency discussion about who provides what services and where the gaps are.

Participants were critical about the lack of program context represented in the evaluation; it was thought to be problematic to treat the program as if it was delivered across sites and across clients in the same way. There were substantial differences across sites, particularly considering the spectrum of need among clients, and differences in the delivery of services: “...it's called Brighter Futures, but effectively in each region it's dissimilar to the neighbouring region. No two are alike” (Earl, BF Managers Group). The manager of one site expressed concern at the lack of structure for the program “it was quite, not confronting but it was a bit, oh wow, is anyone going to be telling us what we have to do here, or what, so it was a bit of a shock really, coming into it” (Lauren, BF Managers Group). This lack of consistency in the delivery of the program was seen to be problematic in evaluating effectiveness at a program level.

One participant highlighted that the evaluation assumed that the practitioners caused change among clients, whereas participants believed that performance in the program was really up to the families; practitioner effectiveness could only do so much. Participants reported that the evaluation read as though the program was operating in the best possible circumstances, even though it rarely was, particularly
in a context complicated by recruitment problems and the challenges of regional service delivery:

...we evaluate as if we’re working in the best possible circumstances when we’re working in the circumstances we’re working in... sometimes we evaluate the parenting program that you would run if the parents didn’t have a drug or alcohol problem (Vanessa, BF Managers Group).

The focus on the prescriptive elements of casework also meant that sometimes the informal work that enables the other elements (e.g. rapport building, lead-in work) was not valued.

While programs like Brighter Futures come with an evidence base and developed rationale, practitioners were concerned with being able to explain the program and have it work for the individual client.

... it needs to have a practical delivery, to make sense to them (clients) and have purpose, and I think that’s where we get off track with some of the things Community Services and Benevolent Society endorse, and they forget that at the end of the day it doesn’t matter what is evidence-based, if it’s working for the parent then that’s what they want to know (Stephanie, Team Leader).

One of the assumptions made by the evaluation was that Community Services and the lead agencies worked with different client groups, yet participants pointed out that this conclusion was based on pre-assessment vulnerabilities "the community referrals that you get, they come with an initially different message, but a post-assessment they often, the assessment often bubbles up the same issues that the helpline referrals get through Community Services" (Megan, BF Managers Group).

Across the participants it was understood that referrals from the community tended to have similar issues to the referrals from Community Services, just that these issues would not present immediately. The conclusion that Community Services staff were more qualified also contradicted the sense within The Benevolent Society that they had more experienced and qualified practitioners than the other lead agencies, and perhaps even Community Services. However, participants also acknowledged that
some sites had significant difficulties retaining experienced staff “... we had recruitment struggles in regional Brighter Futures, but still I don’t think we watered down the uni requirements” (Earl, BF Managers Group).

Another issue with the credibility of the evaluation was that the time delay between the data collection and the dissemination of the research was considerable. With the evaluation taking four years, considerable changes in practices and program design had occurred. Practitioners felt the evaluation findings reflected the program in a different time and had limited meaning in the new context. The problems with the quality of the data also challenged the use of evaluation data for practice "the on-the-ground personnel were quite dubious about [the evaluation data]... those that understand data" (Earl, BF Managers Group).

Among groups in The Benevolent Society, there was a sense that because of problems with the quality of the data it was a bad decision to buy the rights to it, and that this had occurred because the research and evaluation team had their credibility invested in the data by then:

I think it became difficult for the organisation having spent money on the SPRC internally to not justify that, and so our social policy research has had a vested interest in being able to use the data rather than bin it (Earl, BF Managers Group).

- Invasiveness of evaluation

There was some tension with the evaluation affecting practice, some parts of practice had to be routinised and standardised for the evaluation:

...you can only name it a structured home visit if it’s a manualised, sequential, home visiting, parenting model. So they were getting really low numbers and going what’s this about, well when you’re going into a home and doing a manualised, structured, home visit... you sort of shoot yourself in the foot here guys, so there have been some challenges over the definitional stuff around that (Megan, BF Managers Group).
Others reported that the standards and definitions used for the evaluation were simply ignored, meaning that some program work was not recorded in the evaluation data.

To a degree, the evaluation was seen to undermine the processes and measures staff were engaged in to deal with service level issues and they were critical of the fact that the evaluation ignored tacit and intuitive knowledge grounded in the delivery of the program:

...it can come from a practice level as much as anything about how I engage families in a better way that supports them staying in the program for a period of time (Lauren, Regional Manager).

The role of evidence based practice in the marginalisation of practice knowledge was a direct issue in one of the regional sites where staff were resistant to delivering an externally developed evidence based program. These staff wanted to be able to develop their own programs based on the observed needs of the community: "...staff saying you know they’re saying they’re (the program) too elitist, too middle class, they’re too difficult, Aboriginal people don’t learn like that, they don’t learn in groups, so we had all this push back" (Lauren, BF Managers Group). The manager of this site described the challenge of having to bring these regional staff back in line with the values of the organisation. She described her staff as primarily from a family support background and less likely to have tertiary level education than staff in metro centres because of difficulties recruiting in the region. Ultimately, staff that could not practice within the framework of The Benevolent Society left over time.

The length and language of the survey was also thought to be problematic, particularly in Culturally and Linguistically Diverse (CaLD) sites "that was quite difficult to... it took a lot of time to get the information from the families" (Christine, Family Worker). At one point Community Services had suggested that families
without the required English to complete the survey could get their children to help with the translation. This was seen as reflecting the fact that Community Services were totally out of touch with the context of practice. Practitioners also reported difficulty in explaining the purpose of the evaluation to the families involved, causing difficulty in overcoming the resistance of some families to complete the survey.

The practitioners were required to administer the survey. Participants thought some of the discussions required to complete the survey with families could be useful to do; allowing families to reflect on improvements even though the survey responses could not be used for case work. The respondents reported families were generally receptive to the survey, with some understanding that it was being undertaken to see if the program was working. There was some concern about the use of caseworkers to administer the survey beyond the possible bias, that seeking this kind of information could affect the relationship with families. While the survey was offered to all families, the needs of the clients tended to be put before the survey, so it would not be administered if a practitioner thought it would detract from the family’s engagement "I would always put the client's needs over the survey's needs, if I thought it would detract from the relationship I would either not do it... or let them fill it out and bring it back to me" (Christine, Family Worker). Particularly in the Bankstown site, one participant observed that the survey was not sensitive to the experiences of migrants, who given their background were quite concerned about the idea of a government seeking information about their lives:

...there was a bit of a strong response from the crew here in South West because of the cultural diversity I spoke about before, people who have come from war-torn countries, and countries where government is someone to be feared might not necessarily want to fill in a document that goes straight to government (Megan, BF Managers Group).
• **Management of process**

Respondents were also inclined to devalue the evaluation as a true comment on their practice because they believed that the way the evaluation was conducted reflected the uneven power relationship between the different organisations involved. Participants reported that from their perspective the relationship between Community Services and SPRC seemed to be quite fraught, challenged by sudden changes in the evaluation, extra requirements, or changes in approach. The respondents reported plans to conduct a controlled trial, which was abandoned when it was realised that the waiting list that was to be drawn on as a control group was too small to be of use.

At another point there was a discussion about conducting a follow up survey:

> ...there was talk of doing a follow-up survey... we asked, well who's doing those interviews, and they said well you are, the lead agencies are doing them, and it was like no-one's told us to do it, and we've no contact with these families once they leave the program. It was even to the extent where they hadn't seemed to have thought through at all how it was going to work, they hadn't even communicated to the lead agencies that we would need to do these interviews... (Victor, BF Managers Group).

There were also issues with who was responsible for parts of the evaluation:

> ...it got into a bit of a mess about who's actually driving this, and the training was very much, yeah so SPRC was involved in it, but it was very much like you must do this, and this is how it will work... (Victor, BF Managers Group).

Early in the program, the evaluators provided training in the use of the family survey. While there was room for discussion, there was no real consultation on the design; issues were often brought up at lead agency forums, but these were never adequately dealt with “…I was part of a lead agency forum that was held throughout the whole period and there was constant issues raised across lead agencies about the evaluation problems, some of which were inherent in the design” (Earl, BF Managers Group). While staff at The Benevolent Society were involved in a lot of discussions and consultation on the design of the evaluation, participants reported feeling that much of their input was ignored.
There had been direct efforts from Community Services and SPRC to increase the number of completed surveys. This became difficult when due to industrial action there was a Public Service Association ban on Community Services staff completing the survey “… they actually refused to use it, there was a period where our staff, lead agency staff were being forced to do the survey, but Community Service staff, weren’t, and that was a bit contentious…” (Victor, BF Managers Group). For participants this served to reinforce the position of the NGO in having to meet the demands of Community Services.

After an extended period of time where the lead agencies were waiting for the launch of a shared statistics portal, The Benevolent Society entered into protracted negotiations with SPRC to gain access to site specific data. There seemed to be problems within Community Services with the idea of The Benevolent Society spending funds on this data that had come with the contract for the program:

…the irony was that we had to kind of hide that (hiring SPRC to extract site specific data) because they (Community Services) felt that it was up to them to approve whether or not good idea, you know you should be applauding an NGO for going wow we’re going to utilise the investment that’s already there and being made in SPRC... at various times they’ve (Community Services) pulled back saying we think we own that data (Vanessa, BF Managers Group).

Within The Benevolent Society there was a sense that it was the organisation’s money and it could be spent however it wanted, particularly given it was being spent on program improvement which Community Services should have been supporting.

Despite the many reservations expressed about the evaluation some of the managers expressed disappointment that it did not continue “…and I don’t think we’ve seen the payoff of the evaluation, so I would have like to have seen it for ten years, fifteen years, and I think you’d really have staff embedding it in their practice all the way” (Megan, BF Managers Group). In contrast, some practitioners were
happy to see the end of having to administer the family survey, as it was seen as an interruption to their real work.

- **Dissemination of the evaluation**

Despite the efforts the evaluators claimed they made to disseminate the practice implications of the evaluation, practitioners still felt as though there were lengthy reports to read and process, and that they were not directly involved in the evaluation. While participants with high levels of responsibility and interaction with The Benevolent Society head office reported being well informed about the evaluation, practitioners at the service delivery end reported knowing very little “...not informed at all really, it’s only since being in a managerial role... I’ve been more exposed to the results of the evaluation” (Monica, Team Leader). Practitioners talked about being aware of PIPs, and knew that an evaluation was occurring, but there was a real lack of ground level knowledge of what was going on with the evaluation and what the findings were "I think people were trying to give us the feedback, but it was difficult to know what it really meant for the service, I don't know if that was really clear” (Christine, Family Worker).

There was also criticism of limited dissemination and awareness of evaluation to the practice level "I actually had no feedback... it would be up to me to search out" (Monica, Team Leader). Across participants engaged in direct practice there was quite a low level of awareness of the findings of the evaluation. It was understood that the evaluation was generally positive, that families were benefiting from the program and the services were operating as expected, but beyond this there was little detail about what else the evaluation could reveal about the program and what feedback it might offer to practitioners.
• **Good practice in evaluation**

Practitioners discussed how they judged practice effectiveness in the context of the Brighter Futures evaluation. While evaluation and performance measurement data was important, effectiveness was primarily judged by informal feedback from clients and supervisors "we have regular what, we call learning circles, where we get together as a team and provide training and staff development" (Miranda, Team Leader). Practitioners talked about the importance of using the correct language with clients, putting personal values aside, being able to reflect critically, directly asking clients about changes, and obtaining feedback from clinical supervision and external practice consultants. Practitioners also highlighted the importance of formal program outcomes measures of children’s wellbeing, parental capacity, and children’s risk of harm, as well as assessments of the home-life of families, attachment, observations, and getting a sense of the child’s experience, to a meaningful evaluation.

Overwhelmingly participants talked about the need for evaluation to be embedded in the routines of practice; less reliant on centralised evaluation and with more capacity for evaluation to be part of everyday practice. There was a sense that evaluation should be used to address day-to-day questions and to assist with the ongoing implementation of programs:

...so what are you going to do differently and how will I know. and how will your client know that this is what you feel, how are they, how are you going to walk away and say I’m a better practitioner because of this and I think that is probably the next layer that we really not at yet (Megan, BF Managers Group).

However, it was recognised that practice takes place in an inherently chaotic context, and that practice level change may reflect the development of competencies through refining work practices. It was also thought that evaluation results should be more accessible to service users.
Practitioners talked about wanting more engagement in the planning and implementation of evaluation. Overall there was a sense that the evaluation did not deal with practice level issues and was more concerned with big picture structural/policy issues. Practitioners wanted evaluation to deal with critical and responsive issues, like what could be achieved with the workforce available, and definitive answers to matters of practice, such as what are the best interventions, and the best surveys and tools to use. They wanted objective practice oriented measures they could use themselves; there was a desire to add evaluation to practitioners’ toolboxes and have it embedded in practice. There was also a demand for more qualitative detail in the evaluation, the kind of information that is directly comparable to practice. There was demand for this kind of qualitative data in the form of direct feedback from families, reasons for exiting, and what parts of the program families value and why. Practitioners also wanted simplified information from evaluation, such as program strengths, areas for improvement, and how these translate to specific areas of work.

It was observed that practitioners tend to be looking for a sense of feeling good about the program from evaluation, and a sense of whether things were working at the local client level. They tended to be responsive to clients, so dissemination needed to be clear about the connection to client outcomes. It was also observed that the professional background mattered a lot in the dissemination of evaluation information, that among different disciplines there was a need not only to target their areas of practice, but also express evidence in a way that was persuasive to that discipline.
5.3 The Evaluation of the St Marys Child and Family Network delivered by Uniting-Care Burnside

As explained previously this case study did not proceed in the same way as the Brighter Futures study; it involved fewer participants, covered a shorter period of time, and importantly did not cover the period following the dissemination of the final evaluation report. While a follow up was always intended to attempt to capture what happened after the release of the final report, this had still not occurred six months after the evaluation was completed. It became clear over time that this was not going to fit into the timescale for the completion for this thesis. On reflection, while this case was not directly analogous to the other, it was complimentary. My interviews with the human service practitioners at St Marys were done over two days and reflected their experiences at the mid-way point of an evaluation after the release of two interim reports. The process of the evaluation and their experiences with it were fresh in their minds. This was in keeping with the approach taken by much of the research into *evaluation influence* (Herbert, 2011). Participants were able to talk about the dissemination and the impact of the phase one and two reports. I also undertook an interview with the research manager who negotiated the participation of the members of the St Marys staff. This served as the end point of the case study period.

5.3.1 Case Study of the St Marys evaluation

Uniting-Care is a national organisation acting as a representative of the Australian Uniting Church. Subsumed within the NSW branch of Uniting-Care Children, Young
People and Families is Burnside, which states in its annual report that its approach is guided by: connecting and engaging with clients; caring for staff; working collaboratively; and seeking to innovate (Uniting-Care, 2011). Burnside run a variety of programs and services across NSW and the ACT, which includes New Parent and Infant Network (NEWPIN) and the programs that come under the St Marys Child and Family Network. NEWPIN St Marys and the other programs align with Burnside’s aims around generational change, and working with isolated families and young parents (Uniting-Care, 2011).

The St Marys Child and Family Network was an integrated service model, drawing on a number of programs aimed at different levels of service need. All the services were based at a single site within a community with high levels of social disadvantage. Along with the literature suggesting that integrated service models appear to be optimal for addressing social disadvantage (Marsh, Ryan, Choi, & Testa, 2006), the expectation was that having an integrated service would make it possible to foster a sense of community and belonging amongst clients. The centrepiece of the network - as the most intensive and expensive (per client) program - was NEWPIN. The program is an "...intensive child protection and parent education program that works therapeutically with families under stress to break the cycle of destructive family behaviour and enhance parent-child relationships" (NEWPIN, 2013, p. 1). NEWPIN was licensed to Burnside from where it was developed in the UK, and was run across six sites in NSW, with a number of additional sites in Western Australia, the ACT, Tasmania, and Victoria. The St Marys site was different from other NEWPIN sites as complementary programs were run at the same site using many of the same staff; the Young Parents Network and

---

8 St Marys is a suburb in outer Western Sydney (spelt without an apostrophe).
the Growing Together supported playgroup were delivered alongside NEWPIN, along with a number of other smaller services. The Community Connector role was a key part of the integrated model in St Marys NEWPIN. This person worked across the programs in the network and built awareness of the network in the local community.

The NEWPIN model was brought in by Burnside in order to address intergenerational disadvantage in the area. NEWPIN offered an intensive and long-term service, dealing with clients who had complex needs, often involving Community Services. Within the St Marys Network is Growing Together, a supported playgroup program that emphasised working with parents from CaLD communities to address issues of social isolation. Along with the supported playground, the network offered parenting programs, onsite childcare, and a fathers group. The network covered the spectrum of needs, from parents on the cusp of child protection, to families at risk of social isolation. Referrals to the network came from a variety of sources including: Community Services, Community Health, Nepean Youth Accommodation Service, Mission Australia, Police, other programs, and self-referrals. Connection to and across these other services was one of the key elements of the network approach.

The principal idea of the integrated service was to be able to refer from within, to be able to bring clients into the network and then have a discussion about what programs they best fit into. It included assisting women to access other services or supporting them through court and child protection proceedings. The underpinning philosophy maintained that having the clients at a single site allowed them to get comfortable with the space and with the staff. A further benefit of having a range of services at the site meant that the greater number of clients enabled the site to offer
transport by a mini-bus. Informal aspects of the centre included allowing time before sessions for the mothers to have a cup of tea and a chat. The ability to be responsive and provide specific and tailored services to clients was also thought to be an important part of the network.

While on the more intense end of the spectrum of needs, the objectives of St Marys NEWPIN ranged from restoration of children and developing parental capacity through to issues of social isolation and building a sense of community. NEWPIN was an expensive program to deliver, so the co-location of services was in part about attempting to have a wider impact from what would otherwise have been an intensive service involving only a few parents. A number of other programs were co-located with NEWPIN. The Young Parents Network was about establishing a sense of belonging and acceptance for families, providing them somewhere they could feel valued as a parent. The Fathers’ group was about responding to a service gap in fostering attachment and men’s role in parenting.

5.3.1.1 The evaluation of the St Marys Child and Family Network

The St Marys evaluation was contracted out to a university partner in order to develop the working relationship for a longer term research grant. It focused on the network, specifically the effects of integrating a number of programs at a single site. The evaluation had both a formative and summative element in wanting to explore the efficacy of the integrated service in terms of child and family outcomes, but also in exploring the experiences of working in this model. The evaluation identified aims around supporting the expansion and development of integrated services, and informing social policy research.

The evaluation involved a mixed methods approach, attempting to capture both "processes of development and change as well as to quantify anticipated and
unanticipated elements of change over time” (Document B\(^9\), p. 6). The evaluation proposed to explore the program’s impact on child well-being and development, and parent well-being, satisfaction, and use of services (the measures are explained in detail within Document B). It also set out to explore the experiences of service workers in the integrated service model. Drawing on established tools the quantitative measures used by the evaluators examined family demographics, child temperament, child development status, and child literacy. The qualitative elements involved semi-structured interviews with parents on their experience of the integrated service and the effects of the program on their parenting. It also included interviews with service workers about their experience of the integrated service model. Observations of the programs functioning were also undertaken by the evaluators. The first phase of the evaluation was undertaken over eight months, starting in August 2008. The second phase began in May 2009, with the final report completed just after the period for this case study in August 2012.

The findings as reported in the first two interim reports suggested that as a service model the integrated approach was working well, and that there were significant benefits from having services co-located. The evaluators found parents and workers both endorsed the value of integration and the sense of connection and belonging that was developed. Significant shifts in terms of clients’ reported capacity to parent were also found, and the parents felt supported by their connection to the network. The evaluation found that the Community Connector role was critical to the success of the integrated approach, particularly in being visible in the community and promoting the service. In terms of child development however, by the end of the second reporting stage no significant change was found.

\(^9\) See Table 4.6 as a reference to the organisational documents cited in this case.
As highlighted at the start of the discussion of this case, there had been limited dissemination at the point I undertook the interviews; the final report had yet to be released. In my final interview, the manager of research and evaluation emphasised that the process of working with the data was about to begin with some discussion about social bonds, expansion of the network approach, and what new networks could learn from the St Marys evaluation. That said some dissemination had been done of the previous phases of the report, the evaluators had travelled to the St Marys site to share some of the early findings. Copies of the report were also available after they had been accepted by the Burnside Manager of Research.

### 5.3.2 Applying evaluation influence to the St Marys evaluation

The Mark and Henry (2004) *evaluation influence* framework was applied to the St Marys evaluation in order to explore whether the mechanisms identified in the framework were apparent and could explain how *evaluation influence* proceeded as a result of the St Marys evaluation. As explained previously, the St Marys evaluation was ongoing at the time this research ended, the two interim reports had been quite influential, culminating in a number of significant collective level changes. As with the previous case study, all of the impacts of the evaluation identified by the organisational documents, and by the participants were collected and analysed using the Mark and Henry (2004) *evaluation influence* framework. The collection of these impacts and the application of *evaluation influence* mechanisms were much more straightforward than in the previous case, as the evaluation was not commissioned by an external party; policy decisions outside of the organisation did not come into play;
the program evaluated was based at one site, and involved a relatively small number of staff. While the evaluation did produce some summative results, it had a significant formative component focusing on how the integrated service model operated. This meant that more of the evaluation information produced was about practice level issues.

Table 5.3 Mechanisms Reported in the St Marys Evaluation

<table>
<thead>
<tr>
<th>Level of Analysis</th>
<th>Mechanisms</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>General</td>
<td><em>Heuristics</em></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• The evaluation confirmed practitioners’ expectations that the network approach was working and that there were benefits from co-locating;</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Provided a sense that the program (and their jobs) was secure and was likely to continue because the program had been evaluated.</td>
</tr>
<tr>
<td></td>
<td>Skill Acquisition</td>
<td>• Improved the capacity of individual practitioners to work with and reflect on evaluation data in their work.</td>
</tr>
<tr>
<td></td>
<td>Salience</td>
<td>• The evaluation caused practitioners to think about the Community Connector role as being front and centre in the network.</td>
</tr>
<tr>
<td>Cognitive / Affective</td>
<td>Opinion/Attitude Valance</td>
<td>• The positive experience of the evaluation process, and the initial feedback of findings led to an openness to work with evaluation;</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Disappointment with not being consulted before the second phase of the evaluation before it was presented resulted in a sense of a betrayal of trust. It reduced their receptiveness to working with evaluation.</td>
</tr>
</tbody>
</table>
5.3.2.1 Individual level mechanisms

Heuristics represents a mechanism of influence where an individual bypasses a process of systematically thinking about information in favour of responding to implicit cues or heuristics (e.g. majority opinion is correct; Chaiken & Maheswaran, 1994). Most of the practitioners talked about the evaluation reinforcing existing expectations that the service was working well, and that there were significant benefits in having an integrated service. The positive findings of the evaluation reports were understood as a kind of validation of the work done in the centre, that effectiveness could be demonstrated. This was encouraging for practitioners and
fostered a sense that the network and their employment was likely to continue “in that shifting economic climate where there’s decisions made about what services are successful and what services might be closed... the fact that we had that opportunity to get evaluated and it showed significant benefits...” (Zoe, St Marys Team Member).

Participants thought that engaging in the evaluation and participating in discussions about the implications for practice developed their individual capacity to work with evaluation, which Mark and Henry (2003) describe as skill acquisition. While there was still a lack of clarity among some participants about the implications for practice, evaluation data being discussed was seen as a positive, and something that had ongoing benefits: “the people I’m working with now have a better understanding of why it was done... that we can see the benefits” (Amy, St Marys Team Member). The evaluation served an important role in developing evaluative capacity, particularly in having practitioners deal with feedback about the service, and their capacity to understand and work with evaluation data. Respondents reported that the dialogue that occurred with the evaluators about how the service operated gave practitioners a sense that if they engaged they would be valued and listened to.

A shift in the importance of an issue is described by Henry and Mark (2003) as salience. The salience of the evaluation at St Marys was apparent where the evaluation spoke directly to the community connector role. It served to highlight to individuals, even to the person in that role, its importance "...as part of the evaluation they talked a lot about my role, I don’t think when I first started the job, I didn’t think it was that important” (Amy, St Marys Team Member). The importance of this role was elevated in the eyes of the participants due to the evaluation, particularly as the community connector worked across all of the programs in the network.
An important mechanism for individual level change is a shift in attitudes Oliver (2008) describes as opinion/attitude valence. From their early participation in the evaluation and the initial feedback of findings, participants reported an overwhelmingly positive experience of the way the evaluation was conducted: “they were very very good, the researchers who came in, they were very good at engaging the mums, they were very very respectful” (Nicky, St Marys Team Member). This had the effect of increasing the willingness of practitioners to cooperate and collaborate with evaluators. This change to attitudes about evaluation was reversed by the disappointment nearly all of the participants expressed about the lack of consultation prior to the presentation of the second phase of the evaluation at a conference:

_I attended a conference and there was a discussion about this wonderful program called St Mary’s Child and Family Network. It was all up on the big screen with the outcomes and what have you and it’s – that’s about me and people I work with and I’ve never seen that information. I guess in some ways that made me a little more hesitant to have a discussion because there was stuff that was up..._ (Zoe, St Marys Team).

The sense of betrayal of trust felt by the participants resulted in a sense of hesitation in working with evaluation.

**5.3.2.2 Collective level change**

The evaluation data had a variety of collective level impacts. Along with being adopted internally, the evaluation formed part of the body of evidence to support funding for positions across other programs "they know that their experiment with the community connector role was something that has been built into a number of other operations, so at an organisation design level" (Lauri, Manager) in both new and existing services. The evaluation entered into policy consideration through the use of data in the preparation of applications for funding new integrated centres. The evaluation was considered persuasive enough to be included in these applications.
arguing for the benefits of integrated services across future NEWPIN sites. *Agenda setting* occurred, and the evaluation itself was attributed with bringing researchers and practitioners together to discuss streams of funding for NEWPIN and different ways of running the program. The evaluation findings were front and centre at these collaborations, with practitioners providing feedback about the evaluation, and what the findings might mean for the way NEWPIN was delivered in new sites.

While the network approach was understood to have worked and there was encouraging results in parental outcomes, the lack of significant changes in the outcomes for children led to some robust discussions about the effect of the program on child development:

*I guess staff were concerned about the children's findings in phase two that they weren't statistically significant... I guess what we were doing as a reference group, as researcher was to go back to the literature and finding out about we don't actually know much about how parent outcomes translate to children, and we really just assume that happens* (Lauri, Manager).

This required a re-examination of the literature and some discussion about how the program could be delivered differently to have a greater impact. In this way the evaluation set the agenda of efforts to think through redeveloping the program. Participants recognised that while there had been limited shifts in the developmental measures of the children, the evaluation had found shifts in the way parents felt about themselves and their relationships with children, and the support they felt in the community.

The evaluation led to *policy oriented learning*, which Oliver (2008) defines as an increased understanding at the policy-making level of the organisation. Participants reported that the evaluation supported the idea that extra value had been obtained from integrating and co-locating services and that this seemed to have had an impact across the organisation, with policies reflecting a greater emphasis on
partnerships and network building. Practitioners themselves identified that the approach was adopted across the organisation, with NEWPIN being integrated with other programs and other organisations. A participant praised the evaluation as laying out the operation of St Marys in a clear way that could be applied to other networks, particularly with questions about how to staff and run programs, which aided the possibility of policy level learning.

A number of other issues brought up by the evaluation became organisational knowledge through policy-oriented learning. Participants identified the fact that the clients using the program really seemed to need the service, that the program was addressing the right client groups: “the parents and children attracted to the program could really use assistance, and that’s always important to know that we’re hitting our target groups pretty well” (Lauri, Manager). In relation to the operation of the network, the evaluation highlighted that despite the many programs offered by a number of providers, and staff that often moved across sites, the network was successful in creating a sense of belonging and place “...it’s quite a mystery to me how they managed to create this sense of belong when really not everybody is on site at the same time ever, there’s a sense of place that they managed to generate” (Lauri, Manager).

While integration had not happened quite as intended, the evaluation reinforced to the participants that the service model worked well and the stakeholders were satisfied with the way the service had been run. The value of being able to move around the network was also emphasised. The evaluation data provided an improved understanding of the local community through the client group engaged in the service. The demographic information allowed the program to be more responsive to the community. The evaluation served as a pilot for a larger research
project collaboration between the evaluators and Uniting-Care, with some of the instruments intended to be used in this study being piloted in the evaluation. Some of the findings were important in planning for this next research project “child outcomes (for relatively high risk children) hadn’t moved, and high risk child are the focus of that study” (Lauri, Manager).

Beyond policy learning, the evaluation resulted in a number of examples of policy change. The evaluation shaped the role of the community connector, in that the evaluation very early on identified that the role was central to the functioning of the network. Originally the sense was that the community connector was important in an outreach capacity early in the life of the network and then should be based in the centre in the later stages, but the evaluation recommended keeping this person out in the community to keep referrals coming in and to raise the profile of the service. Community connectors had been built into other programs and integrated services across the organisation as a result of this evaluation finding "we applied for with a lead agency for an integrated centre in Mt Druitt... and we built a community connector role into that...” (Lauri, Manager). The integrated service model had also been adopted across the organisation, with additional programs being introduced at sites where NEWPIN was already offered.

5.3.3 Participant perceptions of the St Marys evaluation

The participants' responses to the evaluation are presented in this section as a number of sub-categories: the culture of evaluation at St Marys; evaluation purpose;
participant experiences of the evaluation; and how the practitioners thought the evaluation could have fostered practitioner learning.

5.3.3.1 Culture of evaluation at St Marys

Across the organisation, participants talked about the role of research and evaluation in understanding how well a program worked. The organisational values were understood as ensuring that the services delivered were grounded in evidence, which for the practitioners working in these programs provided a sense of reassurance about the likely success in achieving the desired outcomes for families.

Some practitioners understood evidence based practice in terms of accountability to the funders of the program, that for a program to be evidence based and for it to be evaluated meant that it was likely to be continued “evidence based means funding you know” (Amy, Community Connector). Generally evaluation was understood as validating perceptions about the program in a way that has credibility outside the organisation. One participant reflected that while a lot of thought and time is spent on evidence based practice, it is not terribly important in terms of actual practice:

...to me personally, it's not that important. But I think from – in the bigger government thing that we’re working under now, you’ve got to be able to prove what you’re doing is going to make – is going to get the outcomes that you say that you’re going to do (Zoe, St Marys Team Member).

That said, participants also thought that having an evidence base to support practice meant that the organisation was able to show that the program was delivered in line with the best possible practices, and that evaluation could help with making program goals clearer.

Participants also talked about the importance of research and evaluation feeding into program and practice improvement. In particular evaluation was thought to be important in helping to develop programs suited to the needs of the local
community and in dealing with operational questions “I suppose it keeps you alive and fluid in what you’re doing and you know you need to change and adapt to their needs” (Amy, St Marys Team Member). While practitioners were also concerned with more informal criteria such as the continued use of the service, some also indicated that the kinds of outcomes measured in an evaluation were important to them as practitioners. The workers had a variety of levels of experience with evaluation; some were experiencing it for the first time, while others with a more extensive history in the community services sector were familiar with it.

Beyond the use of evaluation tools with clients, the practitioners connected evaluating their own effectiveness with reflective practice that took place at the individual and team levels, and in supervision. The participants felt there was a strong reflective culture in Uniting-Care Burnside and the St Marys Network, and that they were able to have discussions if something was not working well in the team, or in their individual practice:

*I think it's always about having an opportunity to be reflective with peers that work in the same way. That supports the way you practice. Opportunity to attend training and good supervision for reflective practice so that if you're not sure about something you can work through it in that clinical supervision* (Zoe, St Marys Team Member).

They described a team based reflective process guided by the team leader where they could talk about specific cases and advise each other. One-on-one supervision also occurred with both senior members of the team, and with external practice supervisors contracted by the organisation. It was thought that in some cases it was particularly important to get advice and guidance from someone who was not too closely involved in the case. Reflective journals were also used as a mechanism to improve practice.
While some thought that evaluation data had credibility in the practice reflection process, and suggested that discussions about practice using evaluation data occurred, others did not really see the connection between reflective practice and evidence. Primarily the team leader worked directly with the data and then brought the issues identified to discuss within the team “Nicky might have a quick read through it, and ask us to have a look through it and sit and talk about it” (Jessica, St Marys Team). Overall reflective practice in itself was thought to be much more important in understanding effectiveness than external evaluations, primarily because it involved working with parents to understand what they wanted out of the program and to review this with them based on their individual desired outcomes "so I don’t probably really need some person that’s got that understanding to write up a research project to tell me that the work that we’re doing is working, because I can see it. Does that make sense" (Zoe, St Marys Team Member).

5.3.3.2 Evaluation purpose
The practitioners said they knew very little about the timeline of events prior to the evaluation beginning on site. The single manager level participant was able to provide a lot of detail on how and why the evaluation of the network came about. She described the context of the evaluation, which began with discussions around setting up an Australian Research Council (ARC) grant to evaluate the NEWPIN program with a university partner, who had the required expertise in quantitative measures that did not exist in the team. As St Marys was coming online there was a discussion about the appropriate outcomes for this integrated approach which led to the idea to evaluate St Marys as part of a pilot study for the ARC grant. The evaluation could have been done internally, but it was contracted out to the university partner in order to test the working relationship. A reference group was assembled, including members of the research and evaluation team in the
organisation. While there was not any real tension between formative/summative outcomes, there was some work involved in making it clear to the organisation that the evaluation was about looking at the program at a particular scale and context (within a network).

Having an integrated service around the NEWPIN program was understood by the participants as something new with inherent challenges, which required evaluation. Participants understood the evaluation to be primarily formative, in that it focused on what the integration of these services looked like, how they worked, and how space and the community connector worked across services and even organisations “...to see how it worked and how it impacted on NEWPIN and the benefits of having the other programs around NEWPIN... that was probably for me what I wanted to know” (Nicky, St Marys Team Member). Participants hoped the evaluation was going to be useful for both the practice and design of programs within Burnside ” I was just hoping they might come up with something a bit different, that we haven’t tried or, let’s try something else a different way” (Jessica, St Marys Team Member). It was also suggested that the evaluation could serve a purpose in helping to set out the service model outcomes, to help clarify the logic of the network. The evaluation was also used to pilot some of the instruments that would be used in the ARC project.

Participants also talked about having expectations of some summative aspects of the evaluation, that it would test the work that was done across the network. Participants felt there was an interest in the expectations and outcomes for parents, and of developmental outcomes for children, and that these would serve to assess the quality of the integrated model. They talked about the evaluation testing if the program was working, and if the network approach measured up to the standard non-
integrated NEWPIN format "so Burnside was really interested in evaluating how that network actually measured up against the standard NEWPINs" (Zoe, St Marys Team Member).

While participants did not have the sense that the program was under threat if the evaluation did not go well, it was thought to be attached to future funding "I can only hope that it works out to be effective and we get funding and so on" (Susan, St Marys Team Member).

5.3.3.3 Experiences of the evaluation
The participant accounts of how they experienced the evaluation have been separated into three sections. First, participants discussed their concerns about the ability of the evaluation to measure effectiveness and to produce data credible to them. Secondly, some of the issues around the invasiveness of the evaluation were discussed, primarily in having to deal with the impact of children’s developmental scores being sent out to parents. Finally, participants expressed a sense of betrayal from an incident where the evaluation was disseminated prior to having an opportunity to receive feedback from the evaluators.

- Data quality and credibility
The St Marys Network evaluation was seen by the participants as something quite different to how evaluation is normally done in Uniting-Care as it was undertaken by an external group, and attempted to measure client outcomes in a more detailed and nuanced way. A lot of the workers were unclear or dubious about how the kinds of personal growth experienced by program participants could be captured or reduced to quantitative measures “when it comes to measuring how someone changes, or how their self-esteem improves, I don’t think, except through observation, you wouldn’t really be able to measure it that much... ” (Claudia, St Marys Team Member).

Overwhelmingly, participants talked about the evaluation missing the informal and
behavioural impacts on children; they felt the evaluation was not able to capture changes that would be obvious to someone who knew the child or had observed their progress. Participants suggested there was a lack of observation or measures of behavioural change and emotional growth, and a reliance on developmental measures around numeracy and literacy "that kind of statistic stuff is easy to measure... but stuff about how much better did they feel after you spent three months with them – it seems like it's a little bit more difficult to measure" (Zoe, St Marys Team Member).

The point was emphasised that while the evaluation was concerned with measuring the impact of the program, the service as a whole was much more than the delivery of specific programs, it was also about families having a safe place to go, things to do, a cup of tea and an informal chat.

Practitioners reported being more concerned with real life outcomes such as clients returning to work, children returning from foster care, and the completion of program plans:

I’ve seen the change, I’ve seen the mums ring me, I’ve seen them when they come and drop in, I’m lucky, I’ve always said that I’m lucky to have that because I’ve been here that long. To see them come through, to see the mums that I had and Tracie had, now working in NEWPIN, now doing their masters in different degrees where they never thought they’d be able to do it, it’s a wonderful thing, and working (Jessica, St Marys Team Member).

Measures of success primarily came from the client’s own perceptions about how their progress was going; clients calling the service long after their completion of the program and telling staff that things were going well was formative in the practitioners’ sense that the program was working. Any external (outside the practice group) evaluation of practice was thought to need an observational element in order to see the significance of some of these ‘real life’ changes that may not lend themselves well to quantitative measurement.
Participants acknowledged that the early phases of the evaluation did not show much change in child development outcomes, but that there were significant social and behavioural changes. There was a sense that practitioners were attuned to these kinds of informal and behavioural changes and that the evaluators could have worked with them to have these kinds of informal outcomes included in the evaluation. The informal part of the service was also thought to be vital, and something again not adequately represented in the evaluation:

"...you can’t just pinpoint one thing, you know you sit and talk to these people for an hour a day, it’s the whole process it’s from when they walk in the door this therapy starts, from when they walk in the door until they leave" (Susan, St Marys Team Member).

Participants also talked about some of the limitations of the developmental tests used, in particular that these tests require a high level of trust for children to be able to perform at their actual developmental level.

Some of the other perceptions were that the service was still developing, particularly the understanding of the community connector role and the interaction between all the different parts of the network, meaning that an evaluation was premature. It was suggested that the evaluation was heavily geared towards measuring the impact of the NEWPIN program, rather than the outcomes attributable to the other elements of the program, consistent with this evaluation being a pilot for an ARC application. On top of this there was some concern that the measures were not sensitive enough to pick up small incremental changes in children's development.

- **Invasiveness of evaluation**

Some of the rapport built up between the practitioners and the evaluators was challenged by the release of some of the developmental results to the client families, although ultimately it was largely perceived to be a challenge overcome by the participants. Following the completion of the phase one evaluation report, scores on
the developmental instruments were sent out to families, which resulted in some distress among them:

I know for the parents that the first piece of information that went out went to their homes. What they saw was something that said, this child is performing at 25% - that means 75% of the population is better than your child. That’s what parents thought when they read the information (Zoe, St Marys Team Member).

Particularly for families enrolled in NEWPIN the scores were often very low, which in the context of the report sent out was often quite upsetting "..the scoring was quite low, so there were a few very upset parents reading a cold report" (Nicky, St Marys Team Member). The practitioners had to work with the parents to understand some of the measures and their implications, as well as talk about some things they could do to help their child developmentally "...so we sort of worked with them, got the parents in and went through the reports, but I would say that was the only thing that could have been done better” (Nicky, St Marys Team Member).

By the second phase, problems with the parents being distressed about the results on the developmental indexes were fed back to the evaluators, who included some recommendations about how to improve children’s development in the reports:

So that was fed back and they did change that for the second bit that came out. The letters came here and we sat with the families and opened them and talked through that stuff and that was much more successful (Zoe, St Marys Team Member).

While practitioners were still engaged in having to counsel parents about their children's results, they were much more prepared, and did not have the same sense that they had to solve problems caused by the evaluation:

When their children were assessed their families were sent out the information on what happened with the child... one of the mums had got the letter and she came in, and I said this is really good because it says even though your child is having difficulties counting, they’ve also given you some ideas about how to help that child count (Amy, St Marys Team Member).
Dissemination

Some participants were disappointed in the lack of feedback about the evaluation, and thought that dissemination and feedback should have been part of the contract with the evaluators. That said, the participants were satisfied with the early (phase 1) dissemination. The evaluators had come and explained the findings to them. However after the second phase of the research was completed the same type of follow up was not undertaken straight away. Practitioners expressed disappointment that they had not been kept in the loop at this point:

"...because I’m Aboriginal, yeah I find a lot of my people, Aboriginal people are in research projects, people go away with a lot of information that helps sort of platform them on to higher and greater things, but nothing’s ever fed back to the people they get the information from" (Amy, St Marys Team Member).

There was a sense that practice knowledge could have brought value to the evaluation; that their experiences could have been valuable in properly interpreting the findings and developing the connection between the measures and the functioning of the program "...our area of expertise is working with these families and delivering this program... I think it’s useful to have an initial discussion with the people who are doing it perhaps if that’s possible" (Zoe, St Marys Team Member).

The disappointment about not being kept in the loop of the evaluation was thought to have an impact on the potential practice use of the findings:

"If you were looking at how this evaluation shaped practice and how is evaluation useful as a learning and development tool, one of the key factors is the people who were being evaluated within a program that’s been evaluated need to have a really good – they need to be kept in the loop throughout all the stages of it as much as possible" (Zoe, St Marys Team Member).

Many did reflect on the lack of impact, that they had not read the evaluation and that most of their colleagues had not either. One thought there had been impacts but couldn't identify anything in particular: "probably I guess, there were changes,"
but they were quite, they weren’t huge changes [long pause], [laughs], sorry its all, I’m just trying to think of something specifically I could tell you that we did look at changing” (Amy, St Marys Team Member). There was a sense that the kind of change that comes from an evaluation is managed at a higher level in the organisation and fed back to practitioners more indirectly.

One of the critical incidents that affected the way the practitioners thought about the evaluation was the presentation of second phase findings at a conference before presenting them to the service:

I attended a conference and there was a discussion about this wonderful program called St Mary’s Child and Family Network. It was all up on the big screen with the outcomes and what have you and it’s – that’s about me and people I work with and I’ve never seen that information. I guess in some ways that made me a little more hesitant to have a discussion because there was stuff that was up.. (Zoe, St Marys Team Member).

Nearly all of the participants talked about this incident with a sense of betrayal and anger "..like what happened down, down in the city, where it was put up and we didn’t have, we didn’t know nothing about it... that was sort of a shock... I think we all felt that one...” (Jessica, St Marys Team Member). There were two main dimensions to the anger about this incident; first, that they had not been spoken to and consulted about the results before it was presented publicly:

...that’s also upset some staff members, that they haven’t been told anything and then later on, perhaps a little while ago there was stuff put up on a board and they didn’t know it was in front of fifty of their peers (Amy, St Marys Team Member).

But they were also distressed about client information being presented, primarily as it violated their values around the sharing and accessibility of information:

I think because of the way we work and how we value people’s personal information... even though we knew it was going to be used... I certainly wasn’t aware that it was going to go around and be presented as papers by people here before we’d had an opportunity to look at it and to maybe even view that thing beforehand (Zoe, St Marys Team Member)
• Good practice in evaluation

In reflecting on how they would have liked evaluation to have been done, the practitioners emphasised the importance of external evaluators seeing the program in action:

... I often think that, and this is no offence to you, but how can someone sit there and evaluate, we can only tell you what we know, but it's very hard to evaluate a whole service when you don't see a service in action (Jessica, St Marys Team Member).

Many liked the fact that the interviewers spent time in the centre first to build rapport, so people could get comfortable with having them there, but also taking the time to observe and to see how the network functioned "I got them to come out before hand they sat in the centre, they sat you know in informal groups, so when they came into the centre they were quite welcome" (Nicky, St Marys Team Member). While there were some challenges with the logistics of organising space for interviews and such, participants reported they felt they were well informed about what was happening, although they noted that they had little input into it. That the evaluators had done interviews with clients to understand what they valued about the service was thought to be a good step. Workers were also impressed with the ability of the evaluators to work with the families, their knowledge of the NEWPIN program, and their willingness to take on board the perspective of the practitioners. It was mentioned that some of the positive rapport was lost when the original interviewer left the project.

Participants talked about the feedback on the results of the evaluation that they had received over time, and particularly early in the evaluation there was a sense that this was done well. Participants reported that the evaluators came out and presented the findings to them to compliment the release of the phase one report. The evaluators also participated in a planning day, and their input was well received.
Being able to participate in the evaluation through sharing practice and contextual knowledge was a key issue identified by the participants, highlighted by a few critical incidents that soured what had initially been a positive experience.

Practitioners thought that among their colleagues research and evaluation were regarded as important inputs to practice, but that their views and their questions about practice needed to be represented in this type of information. There was a recognised need to engage in a knowledge exchange between practitioners and evaluators; an acknowledgement of the value of practice knowledge alongside research knowledge. The manager of research in the organisation emphasised the importance of drawing on practice wisdom and the implementation of programs to ground evaluation in the realities of program delivery “...evidence isn’t the only part of the story, so practice wisdom is a big thing, so the way my team works is to start with a lot of listening” (Lauri, Manager). She talked about this as an exchange model of evidence, a respectful dialogue between researchers and practitioners about what they did not know, which respected the knowledge of colleagues across the organisation. Practitioners also wanted the views and experiences of clients to be more prominent in evaluation to find out what they were not getting from the service “...what can we provide that will make people come, and when you look at research it gives you an idea of what is lacking” (Amy, St Marys Team Member).

5.4 Chapter Summary

This chapter presents the findings of the thesis in the form of two case studies, which are analysed using Mark and Henry’s (2004) evaluation influence framework. This analysis provides a lens through which the experiences of the participants can be viewed, in order to develop an understanding of the conditions and barriers to practitioner learning from evaluation. Brighter Futures is a large early intervention
program funded by Community Services and delivered by a mix of lead agencies and Community Services branches. The case study focuses on the evaluation of Brighter Futures as delivered by The Benevolent Society. The evaluation of the program primarily drew on the results of a survey designed to measure parental capacity and family function, along with a data set of client demographics and output information. The evaluation was found to be quite influential within The Benevolent Society at an individual, interpersonal, and a collective level. Individuals discussed the evaluation confirming their perceptions of the program, of understanding the connection between their work and the program as a whole, and an increased sense of the positive contribution evaluation could have in improving program practice. Through interpersonal mechanisms the evaluation continued the development of a culture of evaluation among practitioners, managers were expected to use and reflect on the implications of evaluation data in the administration of their sites. The evaluation also affected practice by causing elements of the program to become more standardised, requiring the program services to be delivered in a way practitioners described as being based on a checklist. In the organisation the evaluation fed into a number of collective learning processes resulting in changes to the program and to internal policy. The evaluation also was connected to significant changes to the programs from outside the organisation.

Participant responses to the Brighter Futures evaluation suggest that a number of practitioner concerns about the evaluation may have limited its influence. These included concerns about data quality and credibility, the relevance of the evaluation data to the program context, the invasiveness of the evaluation, management of the process particularly the lack of practitioner consultation, and the lack of effective dissemination. Practitioners discussed wanting more engagement in
the planning and implementation of evaluation, and for evaluation to be embedded in
practice in order to deal with the kinds of issues practitioners are concerned with.

The second case study concerns the evaluation of the St Marys Child and
Family Network, delivered by Uniting-Care Burnside. The evaluation focused on the
novel aspect of the program, the idea of integrating a number of programs at a single
site in order to build a sense of connection and community. The evaluation was
commissioned by Uniting-Care in order to examine how integration had occurred,
through formative as well as summative approaches. At the individual level the
evaluation confirmed practitioners’ expectations about the benefits of the network
approach. It also improved the reported capacity and willingness of practitioners to
work with evaluation, while a specific incident challenged a number of participants’
trust and willingness to work with evaluators. At the organisational level the
evaluation was influential in expanding the network approach to services across a
number of other sites, which included the funding of Community Connector
positions at new sites. The evaluation prompted a discussion about the structure of
the program, whether improvements in child development could be achieved by
improving parental capacity.

The participant responses to the evaluation provide a sense of some of the
factors that may have limited the influence of the evaluation. Practitioners questioned
the data quality and credibility of the evaluation, that the measures could adequately
capture subtle behavioural change. The way the evaluation played out meant that
practitioners had some concerns about the invasiveness of the evaluation, along with
how dissemination had occurred. In particular the presentation of the evaluation
findings at a conference before the information had been fed back challenged the
practitioners’ sense of involvement in the process. Practitioners thought that in the
early stages the evaluators had worked well to build rapport with themselves and clients, but would like evaluation to do more to foster their participation and draw on their practice knowledge.
6. CHAPTER 6: THE CONDITIONS AND BARRIERS TO PRACTITIONER LEARNING FROM EVALUATION

6.1 Introduction

While evaluation is often associated with accountability and the interests of the funders of social services (e.g. Harris, 2003; Meagher & Healy, 2003), evaluation as a trans-discipline has developed a clear commitment to fostering program improvement and learning (Section 2.4.3). The organisations that deliver human services by and large share this commitment to evidence informed practice and the use of evaluation to improve the delivery of programs (Section 2.3). Given this, what are the conditions and barriers to the kind of practice oriented evaluation described by Schwandt (2005)? This chapter will present the conditions and barriers to practitioner learning from evaluation developed from the case studies presented in chapter 5, which leads to the model of evaluation to influence practitioner learning presented in section 6.4.

There are two aspects to this research that make it particularly novel. The evaluation literature has predominantly explored evaluation use in terms of decision-makers operating at the program or organisational level. This study focuses on human service practitioners, arguing that practice learning is essentially an individual process. In section 3.2.4 I drew on Antonacopoulou (2001) to describe learning as “the liberation of knowledge from self-reflection and questioning” (p. 328), arguing that while organisational culture and collective systems of learning can legitimise information, learning ultimately depends on the individual and the result of their personal process of reflection. I have argued that particularly in the realm of human service practice, individuals have significant autonomy in the way they work,
notwithstanding the perceived sense of increasing managerial control (Section 1.7.3). This suggests that while organisations can legitimise and disseminate evaluation information, collective level change or learning may not translate into practice change unless individual practitioners internalise the information. Organisations can legitimise information or guidelines for practice, but these can be ignored, worked around, or complied with in a token fashion (e.g. Brignall & Modell, 2000; Chang, 2006).

The second novel part of this research is the use of evaluation influence as a framework with which to understand the effects an evaluation has had on the organisation, specifically those effects related to practice learning. As outlined in Section 2.4.2, the term evaluation use has been prominent in the literature, however evaluation researchers have suggested that evaluation influence may be a more appropriate and fruitful approach (Section 2.4.3). Evaluation influence was well suited to this research in particular because it includes personal and interpersonal processes and changes that are important for the definition of learning used, evaluation use is mostly descriptive of evaluation for decision-makers (e.g. instrumental use when it changes a decision; conceptual use when it informs the frame of reference for a decision; symbolic use when it supports a pre-supposed decision). In this approach, evaluation influence describes change mechanisms that lead to learning, which may in turn result in practice change.

An important outcome of this thesis beyond the presentation of the conditions and barriers of practitioner learning from evaluation (Section 6.2) is the adaption of these factors into a model of evaluation to influence practitioner learning (Section 6.4). While chapter 7 lays out some broader recommendations, this model presents some practical steps evaluators can take to improve the influence of their work.
6.2 The Conditions and Barriers to Practitioner Learning from Evaluation

From the two case studies a number of broad themes emerged describing the factors that challenged the influence of the evaluation for practitioners, connected to the way the two evaluations were done. Despite the clear organisational commitments to evaluation and evidence based practice, and efforts to support these, there lacked a clear sense of how evaluation is supposed to be incorporated into practice; there was a lack of clarity about the relationship and interaction between evaluation and practice (Section 6.2.1). This theme seemed to be the backdrop for other issues identified. For evaluation to be influential, it needed to provide valid evidence to inform practice (Section 6.2.2). This included not only methodological validity in the sense that researchers understand it, but a kind of validity and credibility linked to the practice context. To engage practitioners in thinking about and using evaluation they need to be included and treated as important stakeholders, not merely as suppliers of the data; equitable and participatory engagement (Section 6.2.3) appeared to be important to the influence of evaluation. With their knowledge of the program as it operates in the context of the community, practitioners can add a lot of value to evaluations of their programs and without their participation evaluators can be seen to be measuring variables that lack a connection to practice:

... the measuring tool (was) about children’s educational capacity, a lot of our children wouldn’t shift in six months time in those developmentally age appropriate cognitive development areas. Not a lot of measurement on the social shift, where that is where we see massive change in a six month period (Zoe, St Marys Team Member).

There is also an important issue in acknowledging the critical politics of evidence (Section 6.2.4), evaluations can be dismissed as just engaging in the institutionalised myth and ceremony of organisations (Meyer & Rowan, 1991); and evaluation can be seen as part of the process of NGOs taking on the functions and
forms of government bureaucracy (Di Maggio & Powell, 1983). Without representing a plurality of views and acknowledging the realities of stakeholders, engagement may be limited. Finally, learning is made difficult by a clash of inter-organisational priorities (Section 6.2.5), which can affect the way evaluation is carried out and the degree of stakeholder engagement. Where different priorities come into conflict, the onus tends to be on the NGO to pick up the pieces and find a way to get the information they find meaningful, while doing what the service funder requires.

*Figure 6.1 Conditions and Barriers to Practitioner Learning from Evaluation*
Figure 6.1 presents the interaction of the three domains of factors reviewed in Section 3.3: evaluation implementation, stakeholder engagement, and the program/organisational context. The factors identified by this research suggest some of the interactions between these different domains, which are described in the following sections (Sections 6.2.2-6.2.5). Underlying these factors is the ambiguity in the relationship between evaluation and practice.

6.2.1 The relationship between evaluation and practice

As described above, the broad difficulty in evaluation influencing practice emerging from this research is the lack of a clear sense of how evaluation and practice are supposed to interact. While organisations and disciplines have commitments to evidence based practice, there is a gap between the rhetoric and how an evaluation finding might actually translate into practice change. Evaluations often do not involve questions relevant to practice (Meagher & Healy, 2003) “...there isn’t an evidence base for a lot of the work we do... when you’re talking about day to day activities that our practitioners do with families, then often there’s very little evidence that you can draw on” (Victor, BF Managers Group), yet practitioners are concerned with their effectiveness and have an interest in information that may help them assist their clients. The challenge this ambiguity presents is central to understanding how evaluation can influence practitioner learning.

A common feature of many evaluations is that they produce findings that are not surprising to people who work within the framework of the program every day, meaning that descriptive information about the performance of a program is unlikely to have significant implications for practice change (Bober & Bartlett, 2004; Weiss, et al., 2005). There is a case for making a distinction between the types of evaluation that produce the kind of big-picture summative data required by service funders, and the applied and operational data that concerns the actions and processes of
individuals. Participants valued both of these types of questions; practitioners wanted definitive answers about the effectiveness of programs, interventions, assessment tools, and approaches, while also wanting evaluation to deal with specific practice level operational issues. Practice change can often be an afterthought, or left for the organisation to manage following on from organisational level change influenced by an evaluation. For an evaluation to live up to Schwandt’s (2005) standard of practice-oriented evaluation, it should engage with practice in an active and embedded way.

Part of the reason traditional evaluation approaches may be limited in influencing practice is that they are more likely to produce information that is actionable and relevant to policy and organisational settings. Even when these evaluations do result in suggested changes to practice, it became clear in the research that they lack the meaning and context of where and how the program is delivered. Practice takes place in an inherently chaotic context, and is often personal and intuitive; participants said that while a practitioner might work with a particular framework in mind, once they step into the room the practitioner has to respond to the client rather than undertake a mechanised routine. In this way developing practice might be better approached by thinking about developing a set of competencies in people rather than a set of routines “...it’s about then how does that translate to competency, so what are you going to do differently... how are you going to walk away and say I’m a better practitioner because of this” (Megan, BF Managers Group).

6.2.2 The validity of evidence to inform practice
It would appear from the data that for evaluation to be internalised and inform practice, it has to be construed as valid by the practitioner. This is consistent with Fleming’s (2011) description of the social psychological processes involved in attitude and behaviour change. Human service practitioners understand the program
frameworks they work within very well, and how they are actually implemented in practice. In order to tell practitioners something they do not know, or something that contradicts their lived experience, a convincing standard of evidence is required. For evidence to be valid to inform practice takes more than validity as it is understood by researchers and evaluators, it also requires a practical connection and a kind of transparency of the logic behind how things are done. While evaluations have to work with metrics and quantitative instruments in summative evaluations, approaches that engage practitioners should avoid being reductive in a way that loses connection with practice. Informal practices enable the delivery of the formal aspects of programs, which can be obscured in evaluation work or even in program models “...it’s the whole process, it’s from when they walk in the door the therapy starts... that’s when the program starts, but on paper the program is just little snippits of that” (Susan, St Marys Team). Evaluation also has to show the connection between measurement tools and the real world observable outcomes that are apparent to practitioners. Some of this can be achieved by working with practitioners on evaluations, and incorporating their practice and program implementation knowledge.

As can be observed in the findings chapter (Section 5.2.3 & Section 5.3.3), the practitioners had quite sophisticated critiques of the evaluations that for them raised questions about the validity of the findings to inform their practice, and the capacity of the evaluation to reflect and measure their work. Across both cases, but particularly in the Brighter Futures evaluation they criticised elements of the design of the evaluation, the way the data was collected, and the findings drawn from the data. An important pre-cursor to influence through the thoughtful mechanisms (i.e. not heuristics) is the positive appraisal of information in terms of its validity and
relevance (Christie, 2007; Fleming, 2011). Practitioners’ knowledge of the program means that the quality of evaluation evidence needs to stand up to practitioner scrutiny, particularly with novel or controversial findings. Of course, evaluations should be rigorous and draw conclusions based on firm evidence generally, but it appears to be particularly important when the intent is to have an impact on practice.

The findings also highlighted how practitioners from different disciplines and performing different roles in the program can vary in their assessments of validity. Many disciplines are still engaged in debate over epistemological issues to do with what constitutes valid evidence (Donaldson & Christie, 2005; Parker, 2002). Scientific professionals can have a demanding standard of validity and reliability in terms of the design of experimental studies and large sample sizes (Christie, 2007). In the Brighter Futures case study, participants observed that staff with an educational background or in more responsive personalised practice roles were dubious about the meaningfulness of surveys and standardised assessment instruments "the tension I found was with the teachers and some of the other helping professions. The teachers had less inclination to use an evidence base than the other professions” (Earl, BF Managers Group). This is consistent with Christie’s (2007) findings in her evaluation influence simulation study that found participants with a background in education tended to be less influenced by survey evidence. The different responses to the data highlight the importance of recognising the discipline background of stakeholders and the types of evidence that are likely to be persuasive to these practitioners.

The data suggests that for evaluations to be meaningful for practitioners it needs to have a connection to the context of practice, acknowledging the realities of practitioners and clients. This is a type of validity beyond methodology; it is about
the connection between the state of affairs described by the evaluation with the experiences of practitioners. Evaluation is inevitably reductive, it attempts to reduce the complexities of human service delivery into common and meaningful metrics (Holland, et al., 2009); evaluations can lack practice validity when they fail to acknowledge this complexity. Across both cases, there was a sense that the evaluation failed to adequately present the delivery of the program; particularly in capturing the informal things practitioners do that support and make the formal parts of the program work. In the Brighter Futures evaluation, one of the main criticisms was that there was a lack of consistency in service delivery across sites; the very different client experiences with the program were presented as equivalent in the evaluation. Often service delivery as it was described in the evaluation bore little resemblance to how practitioners believed it was carried out in practice; practitioners complained of the evaluation dealing with idealised conditions of service delivery rather than actual conditions. On the outcomes side, across both cases participants thought the evaluation had been reductive in the way the outcomes were measured, that they did not connect to, or failed to represent the kinds of observable behavioural outcomes that were important to practitioners.

In both the case studies there was a significant gap between the completion of the data collection and the presentation of the final report. Particularly in The Benevolent Society there was some criticism that the findings related to the program at a previous point in time, that the data lacked practice validity because it did not reflect the way practice was currently being done. The long turnaround time for research and evaluation to occur is an inevitable challenge to practice relevance (Nutley, Walter, & Davies, 2003), yet it also highlights another validity issue that also connects to the politics of evidence. Proceeding with an outcomes evaluation
with a new or developing program is questionable; programs do and should change and evolve over time in response to the practice context. Trochim (2009a) argues that evaluation needs to be used to create “a supporting system of evidence norms and phased trials that provide a necessary foundation for [Randomised Controlled Trials]” (p. 29). That is evaluation should play a role in: (a) developing and articulating the program framework, (b) assuring the consistency and fidelity of implementation, (c) identifying appropriate outcomes measures, (d) developing the capacity for change in the program, and (e) creating the environment for an effective outcome evaluation to be done (Trochim, 2009a). While there is a political imperative to produce evidence about the effectiveness of a program, the meaningfulness of this evidence is limited unless the program is in a stage to be evaluated.

Part of practice validity seems to be a kind of transparency to service users; that the programs have a logic behind them that can be explained in a clear and concise way. Participants favoured evaluation forming part of the accountability of the program to service users, a sentiment that contrasts with research on the managers of NGOs (Baulderstone, 2007). Evaluation should play a part in this by connecting practice to outcomes; that at the casework level the connection between the program and the desired outcomes is made clear.

Validity to inform practice connects directly to stakeholder engagement, if practitioners do not feel the evaluation reflects the work they do, they are unlikely to engage, which may minimise influence. Where practitioners feel that evaluation data is valid and has consistency with their practice experience, engagement may be increased. That said, the relationship between stakeholder engagement and validity can work both ways; working with practitioners and involving them in the design of
an evaluation and the interpretation of the results can add practice validity. By engaging staff who have the most experience of the program in action (Ebrahim, 2005b), drawing in their practice knowledge, and their capacity to create solutions (Arnkil, 2012) including practitioners as stakeholders can add validity and practice relevance to an evaluation.

6.2.3 **Equitable and participatory engagement**

Much of the evaluation literature since the early 2000s has emphasised the importance of fostering the impact of an evaluation, primarily by having a clearly defined role for important stakeholders (Johnson, et al., 2009). Staff who pursue the aims of the program through their daily practice, and attempt to achieve the intended outcomes are vital stakeholders of any evaluation (Ebrahim, 2005b; Naccarella, et al., 2007). When they are not treated as such, evaluation is unlikely to be influential, and may engender indifference or even resistance (Donaldson, Golder, & Scriven, 2002). Effective communication over the course of the evaluation (not just the conclusion of it) is an important part of engagement, connecting the future benefits to any current additional expenditure of time and effort (Bamberger, Rugh, Church, & Fort, 2004; Bober & Bartlett, 2004). Valuing practice and program implementation knowledge can enhance the validity of an evaluation, but it can also serve to engage practitioners and develop their capacity to work with and understand evaluation information. While there is a short term benefit of producing more practice relevant information and improving the influence of the evaluation, the long-term benefit is the ongoing interest and buy-in to the use of evaluation to inform practice (Stevenson, Florin, Mills, & Andrade, 2002). While engagement begins with the design of the evaluation, how this engagement plays out and how it reflects the relative importance of different stakeholders is highly significant. Achieving equitable and participatory engagement may enhance the implementation of an
evaluation through practitioner input on the design of evaluation. The ability to
develop this kind of engagement can depend on the program and organisational
context. This effect may flow back in the other direction also, with the type of
engagement impacting on the context of the program.

An important underlying element of engagement is communication, in
particular effective ongoing communication about the evaluation. While evaluations
may have dissemination plans in place, engagement begins from the planning stage,
communicating and collaborating on the purpose of an evaluation, how it is going to
be undertaken, and how practice effectiveness is being measured. This prolonged
engagement then enables the possibility of an open dialogue about practice relevance
(Thomas & Palfrey, 1996). Communication and dissemination were quite different in
the two case studies. In the Brighter Futures evaluation, The Benevolent Society
had to become active in engaging staff in the evaluation and in communicating what was
happening and why; much of the responsibility for practitioner engagement fell to
the NGO as this was not something that was planned into the evaluation. In the St
Marys evaluation, practitioners initially reported very good communication, rapport,
and respect from the evaluators, although they weren’t in a position to collaborate or
contribute significantly. This was less so at the second phase of the evaluation, where
the rapport broke down after practitioners were not consulted prior to a conference
where the results were presented.

While communication and the rhetoric of collaboration can be engaging, how
it actually occurs in practice is possibly more important. While practitioners can be
acknowledged as significant stakeholders in documents, what happens and how they
are treated in the process of the evaluation is likely to be much more salient to them;
particularly where their advice and suggestions are not addressed. In the Brighter
Futures evaluation practitioners had to administer the family survey which was long and initially paper based. On top of this challenge, there was no practice or client benefit from the survey as the information went off to Community Services with limited feedback. The implementation of the evaluation was inconsistent and the attrition rates were extremely high (Document D). The St Marys evaluation had two events in the evaluation process that affected the engagement of practitioners. The presentation of findings at a conference prior to providing feedback to the practitioner group seriously undermined engagement, with some participants expressing reluctance or apprehensiveness about working with evaluators as a result. Up until the second phase there was a high standard of equitable and participatory engagement set, which was violated by the presentation of information about clients and their practice team without their consultation. Practitioners also talked about having to counsel families when their children’s developmental results were sent out over the course of the evaluation. While they were able to turn this into a positive by including some things parents could do to improve their child’s development in these reports, it fell to the practitioners to resolve the initial distress of the families.

Establishing equitable and participatory engagement with practitioners seems to be important for the influence of evaluation. This type of engagement leads to the successful implementation of evaluations with the kinds of evidence that are valid and meaningful to practitioners, while also building ongoing evaluation capacity in the program/organisation. By properly including practitioners as valued stakeholders in an evaluation, practice and program implementation knowledge can be used not only to enhance the quality of an evaluation, but to embed evaluation into practice through developing capacity. Practice knowledge needs to be valued and considered alongside research and evaluation knowledge in order to influence practice.
6.2.4 The politics of evidence & evaluation

The political implications of evaluation within organisations and the connection to the broader politics of evaluation and evidence may present challenges to the influence of evaluation. Evaluation that fails to acknowledge the political context and critical realities are likely to limit their influence among practitioners (Markiewicz, 2008). Organisations frame evaluation and change within a dominant structure that often does not question implicit or take for granted norms (Antonacopoulou, 2006). Evaluations that attempt to stay neutral to these politics may lack usefulness and relevance (Patton, 1997), and reinforce existing power relations (Meagher & Healy, 2003; Schwandt, 2005). This political element comes into play when an evaluation fails to acknowledge or question adequately a taken-for-granted understanding, a perceived critical reality, or different sides of an issue. This can relate to elements of a program, evaluation, organisation, or even the sector as a whole. Addressing the critical issues of practitioners as experienced in the context of the program may foster engagement, and lead to a better evaluation by questioning the implicit, and representing the views of disempowered and marginalised groups.

An important issue for evaluations intended to foster practice learning is to represent practice realities; a representation of the program as close to possible to how practitioners experience it. Discovering this will inevitably require a process component to the evaluation to explore how the program is implemented. In the Brighter Futures evaluation, participants were concerned that the evaluation seemed to be based on an idealised version of the program, rather than anything resembling the sites they worked in. The program was understood among practitioners to be working with clients far beyond the program specification, this was addressed in the evaluation to some extent. Practitioners were critical of some of the findings that did not represent their experience of the program; particularly the comparison of
Community Services and lead agency service delivery. That said the evaluation helped to acknowledge some important practice realities and begin the process of addressing them.

Along with critical realities of the program, acknowledging and engaging with the political interests of the evaluation itself is important; providing clear answers about what the evaluation was for. While organisations may understand improvement in terms of efficiencies and effectiveness, practitioners may be more oriented towards the social mission of the program and organisation (Harris, 2003; Johnson-Abdelmalik, 2011). Presenting the state of affairs honestly and critically is vital for an evaluation to connect with practitioners and resonate with their sense of the world. For example in the Brighter Futures evaluation, there was a lack of clarity around the connection between the comparison of Community Services and lead agency service delivery. While the government response to the Wood report (Wood, 2008) indicated that the decision about who will run the service would wait for the completion of the evaluation, there was a sense by practitioners that this was a decision that had already been made. The evaluation report itself (Document D) highlighted that the comparison was limited, while participants were dubious about the inclusion of a comparison in the first place. There was some suggestion that Community Services had influence over what went into the report in addition to their influence over the design of the evaluation. The data suggests that evaluations need to acknowledge the politics around the meaning of and use of the findings produced, or risk misuse and cultural irrelevance.

The relationship between evaluation, evidence, and practice has wider socio-political implications, particularly connected to perceptions of human service practice being affected by neo-liberal managerial practices. The role of care workers
(Meagher & Parton, 2004) and the kinds of critical skills and knowledge associated with this work (Baines, 2006; Parton, 2008), are thought to be under threat from the change to a bureaucratised form of social work. The concern is that practice becomes defined by groups distant from clients and the frontlines of the program. In the Brighter Futures evaluation, participants were concerned that the evaluation started adversely affecting practice; that things had to be done in a certain way in order for it to count for the evaluation. They also thought that the Community Services practitioners exhibited less personalised and responsive practice to clients, which was reflective of a risk based approach. The data suggests - consistent with Cousins’ (1995) study of effective participatory evaluations - that evaluations need to engage with the broader context of the issues important to the relevant stakeholders.

Evaluators should endeavour to represent the political implications of programs and evaluations in order to be relevant to the practice context, particularly where there is a sense that some perspectives are marginalised in the organisation. Evaluations that fail to foster and acknowledge practitioner perspectives or are obtuse about the purpose of an evaluation are likely to fail to connect to the context of practice. This means evaluations can be based on assumptions that do not reflect the practice context, which can affect the implementation of the evaluation findings, and the engagement of stakeholders.

6.2.5 Inter-organisational priorities
Interrelated with the issue of politics and evaluation is the clash of inter-organisational priorities and values, particularly those between the funders and providers of services. In a model of social work where funders purchase services (O'Shea, 2007; O'Shea, et al., 2007), governments become increasingly distant from the practical knowledge that comes with delivering the service. This distance may mean that practice and managing practice improvement is seen as the responsibility
of the NGO, and not something that should be included as part of an evaluation designed for the funders’ needs. Particularly where practice improvement is the secondary purpose of an evaluation there may be a lack of clarity over the roles for the different parties involved. The ability to conduct evaluation in a way that engages practitioners can be challenged by inequalities in the interactions between the different organisations; this can be thought about in terms of: (a) what the evaluation is for; (b) how an evaluation is conducted; and (c) how the evaluation is disseminated. While practice improvement represents value-adding for an evaluation (Ebrahim, 2005a), and is clearly in all party’s interests, organisations may fail to recognise the connection between practice improvement and program success. This has implications for the way evaluation is carried out, and for effective stakeholder engagement.

In both the case studies it was clear that practitioner learning was not a priority. In the Brighter Futures evaluation despite The Benevolent Society making an effort to get data and findings they could feed back, and implementing processes to engage practitioners, this was made difficult by the lack of focus on practice in the evaluation. Despite the rhetoric and the inclusion of a process evaluation, the Brighter Futures evaluation appeared to participants to have been conducted to show to the next level of accountability upwards that Community Services had invested well in the program. The types of evidence the evaluation produced reflect evidence that would be compelling to NSW treasury. There is in part an intractable problem of having to conduct summative evaluation before a program is ready for it; Trochim (2009a) outlines a series of evaluation processes that should occur before conducting a large scale evaluation. However, even with the need to produce a judgement of
worth (albeit a premature one) about a program, there are still opportunities to foster practice learning.

While Appleton-Dyer, Clinton, Carswell, & McNeill (2012) present a comprehensive theory on *evaluation influence* in public sector partnerships, the current findings suggest a complimentary, but simpler connection between *evaluation influence* and the interaction of partnered organisations. As the purchasers of human services, government departments may be isolated from the delivery of services and the imperative for practice improvement. As well as the loss of program implementation knowledge, the loss of focus on practice is also likely to be reflected in the way evaluation is done and what it is understood to be for. That is, an evaluation that a government department has influence over is likely to reflect their need for accountability, and it is likely to be undertaken in a way that has legitimacy to their stakeholders (Di Maggio & Powell, 1983; Ebrahim, 2005a).

The Brighter Futures evaluation is a complex example; it was a mixed service model for much of the life of the program. This meant that the service was delivered by the NGOs alongside practitioners in Community Services, and each site had to work with their local branch of Community Services, along with additional partners (e.g. local health services, migrant resource centres). Underlying this complexity, the relationship between The Benevolent Society and Community Services in the Brighter Futures evaluation was characterised as being inequitable (Document D, p. xii). This seems to be reflected in the way the evaluation was carried out: participants reported that consultation with the lead agencies was ignored. For example, additional components of the evaluation were added on with which the lead agencies were expected to comply with, and practitioners were expected to administer the lengthy family survey for the evaluation with no practice benefit. Practice learning or
program improvement was not really part of the agenda for the evaluation, it seems to have been included as an afterthought in the final evaluation report (Document D), rather than embedded in the aims of the process evaluation. Responsibility for program improvement was taken on by The Benevolent Society, who contracted the evaluators to extract site specific data, and put processes in place to feedback information and disseminate key findings.

Without effective coordination, there can be lack of clarity around the roles and responsibilities for the different partners in an evaluation, and in some cases it can be in the interests of parties to obscure some of the responsibilities. Within The Benevolent Society this was experienced as having to take responsibility for the effective dissemination and implementation of the evaluation findings, despite some of this being the usual province of evaluators (Lawrenz, et al., 2007). Despite not having an external funder driving the evaluation, practitioners thought that more dissemination and working with practitioners on the findings should have been included in the contract to evaluate the St Marys Network. In the Brighter Futures evaluation, much of the detail on how the evaluation was going to be conducted was not revealed until after the service contracts were signed. Furthermore the promise of an online statistics portal delayed the organisation’s efforts to collect and work with their own data. By obscuring the full extent of what would be eventually required of the lead agencies; Community Services exploited a position of informational power, which meant that The Benevolent Society had to take on extra responsibilities, costs, and efforts to be able to get value out of the evaluation.

With broader trends removing responsibility for practice from the view of funders of programs and evaluations (Ebrahimi, 2005b), there may be an increasing tendency for evaluation to ignore the prerogative to produce information of relevance.
to practitioners. This is despite contemporary evaluation practice emphasising participatory and collaborative approaches as the way to achieve practice and organisational change (Hall, et al., 2012; Schwandt, 2005). Funders who adopt a myopic view of practice (Ebrahim, 2005a) ignore the value-added nature of practice oriented evaluation, and the benefits it can have to both program functioning and the ongoing development of a culture of evaluation and evidence.

6.2.6 Summary of interrelationships
The model (Figure 6.1) describes how the themes identified from the case studies connect to the categories of influence factors identified in section 3.4. The backdrop for these themes is the relationship between evaluation and practice; each of the themes represent a smaller scale part of the essential problem of the lack of clarity about the relationship, or how exactly evaluation evidence can be incorporated into practice.

The validity of evidence to inform practice represents not only the validity and reliability of evaluation data, but the credibility of that evidence to inform practice. This is an issue connected to stakeholder engagement in two ways: having evidence that is convincing to practitioners is likely to increase engagement, and having proper practitioner engagement in the evaluation is likely to result in data with practice validity.

Conducting an evaluation that recognises practitioners as important stakeholders, and achieving equitable and participatory engagement may be central to promoting the influence of an evaluation in human service practice. It was clear from the research reported here that a failure to adequately engage with practitioners casts the evaluation as ill-informed at best. Engagement relates to the way the evaluation is implemented, that respect and knowledge exchange is demonstrated in the planning and conduct of evaluation (Scarinci, Johnson, Hardy, Marron, &
Partridge, 2009). The opposite also follows, as engagement can assist evaluators collect the kinds of evidence that is valid and meaningful to practitioners (Daigneault & Jacob, 2009). Conducting evaluations that engage well with practitioner can also build ongoing evaluation capacity in the program/organisation that can lay the groundwork for effective engagement in future evaluations.

Acknowledging the politics of evidence in the evaluation, and the broader socio-political issues linked to the interaction between evaluation and practice enables an evaluation to better reflect the perspectives of practitioners (e.g. acknowledging that evaluation is often associated with accountability in order to explain how this evaluation will foster learning). This has the effect of improving the implementation of the evaluation by putting the sometimes implicit or unquestioned ways of doing things on the table. Primarily acknowledging politics is about improving engagement by reflecting the lived experiences of practitioners as stakeholders in the evaluation, conversely effective engagement results in the politics of evidence being dealt with in the evaluation.

Finally, the recognition of different inter-organisational priorities in the evaluation and the problem of an unclear set of roles and responsibilities for the different organisations involved directly affects the way the evaluation is implemented; an evaluation can be carried out poorly if there is confusion about who is supposed to do what, or if there is ongoing conflict about the different priorities of the evaluation. Within the Brighter Futures evaluation, the ambiguous arrangements between the different agencies led to occasional conflicts about which agency was responsible for parts of the evaluation. Conflict can affect engagement, such as where one of the organisations is unable to get practice improvement included as a
priority for an evaluation. Effective engagement of practitioners in the evaluation can enable them to advocate for practice improvement as a priority for evaluation.

6.3 **Comparison with the Factors in the Literature Associated with Evaluation Use**

Section 3.3 included a review of studies that explore the factors associated with evaluation use, which were condensed into a set of factors to explore in this study (Figures 3.1-3.3). These included factors to do with the way the evaluation was implemented (Cousins & Leithwood, 1986; Grob, 2003; Johnson, et al., 2009), the way stakeholder involvement was managed (Diaz-Puente, et al., 2009; Diaz-Puente, et al., 2008; Johnson, et al., 2009; Shulha & Cousins, 1997; Vanlandingham, 2011), and program or organisational factors (Balthasar, 2009; Cousins & Leithwood, 1986; Johnson, et al., 2009; Karan, 2009; Murphy, 2007). These categories of factors have been used as the frame within which the conditions and barriers are considered in this research. The research these factors were drawn from mostly involve studies of decision-makers or policy-makers, drawing on evidence of how evaluation affected their decision-making in order to make inferences or demonstrate the significance of particular factors. For the present research I have argued that the practice context is not analogous to decision-making or policy-making, practice is not adequately represented as a set of micro-decisions or procedures (Evans, 2011; Evans & Harris, 2004). In fact some would argue that such procedural work is an anathema to ethical and caring human service practice (Meagher & Parton, 2004). As practitioner learning represents a substantially different phenomena from what has been explored in the literature so far, the factors drawn from both the use and influence literature were not directly used in the analysis. This section serves to compare and contrast the factors found in this research with the factors in the literature.
The factors identified in the literature (Section 3.4) as being associated with the use of evaluation (primarily use by decision-makers or policy makers) include the following:

- Evaluation Implementation: Quality of the evaluation, relevance, credibility, findings, timeliness, communication, competence;
- Stakeholder Involvement: Communication quality, direct stakeholder involvement, credibility, findings, relevance, personal characteristics, commitment or receptiveness to evaluation, decision characteristics, information needs;
- Program or Organisational Setting: Competing information, personal characteristics, commitment or receptiveness to evaluation, program characteristics, information needs, political climate, organisational learning, leadership.

Compared to these factors, the conditions and barriers drawn from the case studies are more dynamic, for example equitable and participatory engagement reflects the accounts of how practitioners would like to engage with evaluation, and how this engagement can be undermined. As the data drawn on are participants’ perceptions of and experiences of evaluation, rather than outside observations of many evaluations, this research cannot make inferences about the contribution of factors as detailed as Johnson et al (2009). The strength of this research is in developing a model of evaluation practice to enhance practitioner learning drawing on the perspective of practitioners and their experiences in having their practice evaluated.

In the evaluation implementation category, this research found a theme describing practice validity; which seems to be a mix of evaluation quality, relevance, credibility, findings and timeliness. Practitioners described a need for not only technical quality and credibility, but for practice relevance and a clear connection between the procedures used and the practice context. Findings suggest
that evaluation that challenges the expectations and experiential knowledge of practitioners need to have robust practice validity in order to have influence. Timing also connects to the idea of practice validity in that evaluation findings should reflect the current practice context to be valid to inform practice.

In terms of stakeholder involvement, *equitable and participatory engagement* describes the fostering of a sustained stakeholder relationship with practitioners, using their practice knowledge to enhance the implementation of the evaluation, to foster the influence of findings, and to build ongoing evaluation capacity. This seems to include elements of commitment or receptiveness to evaluation, communication quality, direct stakeholder involvement, credibility, findings, relevance, and information needs. By involving practitioners as valued stakeholders in an evaluation, their information needs and context (related to relevance) can be built into the evaluation, which translates to improved credibility and knowledge of findings. Direct involvement in the evaluation also translates to improved influence, along with an increased receptiveness to evaluation.

The *politics of evidence* and *inter-organisational priorities* relate to the organisational setting, each describing a challenge to creating the environment for evaluation to be conducted well and to influence practice. The *politics of evidence* describes the need to acknowledge the realities of the political climate; of the information needs, in the sense of whose information the evaluation is for; and the practice characteristics. This also encapsulates competing information in the sense that evaluations need to acknowledge the existence of other information and perspectives. *Inter-organisational priorities* concern the political climate around the management of evaluations, but also the leadership and organisational processes
around evaluation. This connects to the organisational or program level commitment or receptiveness to evaluation.

6.4 A Model of Evaluation to Influence Practitioner Learning

Incorporating the themes drawn from the case studies, a clear picture emerges of an approach to evaluation that develops practitioners’ capacity to think about and use evaluation information in their practice. Acknowledging the importance of evaluation influence as a framework to understand processes of change, this model sets out an approach to evaluation that aims for practice change achieved through practitioner learning. This approach deals with the critical issue of the relationship between evaluation and practice by:

- Embedding evaluation into practice: Using evaluation information as a tool for practitioners to understand and reflect on their effectiveness in practice complimenting its role in producing evidence on program level effectiveness;
- Creating evidence that is valid to inform practice: Ensuring that practice is evaluated in a way that is methodologically valid, but also retains a clear link to the practice perspective;
- Making effective use of practitioners as evaluation stakeholders: Properly valuing practitioners as stakeholders, incorporating their program and implementation knowledge, and treating them ethically and equitably;
- Engaging with the critical politics of organisations and the sector: Ensuring evaluations question implicit and taken for granted assumptions and avoid marginalising particular groups; and
- Advocating for practice relevant evaluation: Communicating the mutual interest in producing findings oriented to practice level learning.

6.4.1 Embedding evaluation into practice

It would appear from this research that one way to increase the practice relevance of evaluation is to build capacity and resources around practitioners participating in and administering evaluation that is complimentary to practice. In the findings,
participants talked about wanting evaluation to be part of their tool kit, to be embedded in practice as a measure of individual efficacy alongside the other means by which they determine their efficacy as practitioners (e.g. reflection, supervision, team reflection, discussions with clients). While seeing the value in program level evaluation conducted by professionals, there was a desire for individually oriented data to help reflect on practice. Importantly this would need to be provided in a casework setting in a way that is not used by the organisation to make judgements about individual practitioners. While the possible extra work involved for individuals, and developing the competencies to be able to draw on evaluation data in case work may present a challenge, there is potential for significant benefits to enrich practice with evaluation evidence. Program level evaluations that make this information available to practitioners also represent significantly improved value.

6.4.2 Creating evidence that is valid to inform practice
While validity and reliability are criteria researchers and evaluators are familiar with, evaluations that are intended to inform practice need to pass another hurdle, practice relevance. Evaluations should consult with practitioners to get a sense of their assessments of validity and reliability, which is informed by their experience of implementing the program, along with their sense that the evaluation connects to the practice context. This connection concerns the level of reductiveness in the evaluation; does the way the program is described and the way outcomes are measured reflect what actually happens in practice. This consultation also needs to recognise some of the disciplinary differences in understanding research and evaluation evidence, that different groups will interpret data in different ways. To be effective in fostering practice learning, evaluators should consult with and understand the groups the data is being produced to inform.
Along with consulting practitioner stakeholders on the practice validity of data, evaluators should display an openness to the practice context, and consider undertaking observations of the program to ensure their conceptualisation matches the daily realities of human service practice. Much human service practice is responsive and informal, and reducing programs down to component parts often misses the full picture of what the program actually does. Attempting to capture the informal work, and the informal outcomes of the program is an important part of ensuring that practitioners are confident that the way process and outcomes have been understood reflect the way they are understood by staff working in the program. Doing this requires a strong process component to an evaluation; attempting to understand how the program works before moving on to measuring its efficacy. This is particularly important in programs delivered across multiple sites that may have significant variations in delivery.

### 6.4.3 Making effective use of practitioners as evaluation stakeholders

To influence practice, practitioners need to be included as important stakeholders of the evaluation. The findings of this study suggest that beyond merely informing them of the progress of the evaluation, practitioners should be engaged as partners, with their input valued on the design and conduct of the evaluation, and the interpretation of results. In order to learn from evaluation practitioners should be treated fairly and equitably over the course of the evaluation and have their concerns and feedback valued and addressed. Engaging practitioners requires an open dialogue about what is being done and why.

### 6.4.4 Engaging with critical politics

Evaluation can often be seen as something ceremonial, or done for funding, accountability or some other concern divorced from practice. For an evaluation to be taken seriously by practitioners, the critical politics associated with the program and
the evaluation should be acknowledged. Evaluations concerned with practitioner learning need to attempt to represent a plurality of views, and challenge the implicit assumptions of organisations. Canvassing and including practice level perspectives about how the program works and what it is for, alongside legitimised and documented perspectives is important in getting towards the truth of a program. The reasons and motives for an evaluation need to be expressed transparently, and reflect perspectives from a variety of stakeholders. In producing evaluation intended to influence practice, there is also a need to address broader issues and concerns about the connection between evaluation and managerialism, and the critical politics of evidence based practice (Meagher & Parton, 2004).

6.4.5 Advocating for practice relevant evaluation
NGOs may lack the power to effectively advocate for practice relevant evaluation where program funders have direct influence over the direction of an evaluation, but need to remain open to opportunities to promote this type of evaluation practice. Evaluators potentially have a role in advocating for evaluation that is practice oriented and to put in place processes that appropriately engage and feed information back to practitioners. Involving and informing practitioners about evaluation findings adds value to an evaluation and potentially increases influence.

6.5 The Utility of Evaluation Influence in Describing Practitioner Learning
This research is novel in that much evaluation research literature employs the concept of evaluation use, and focuses on the impact of evaluation on decision-makers and policy-makers. This section will provide some reflection on the significance of the evaluation influence framework (Mark and Henry, 2004) in this research, and of the significance of investigating practitioner learning in the factors that seem to be important for influence to occur.
The application of *evaluation influence* to the data was undertaken with the benefit of a full review of studies that had also attempted to identify influence mechanisms in retrospective case studies (Section 2.7). The review identified a number of limitations in the development of *evaluation influence* as a research approach. First, the review identified that much of the research has been undertaken by individuals with a clear stake in presenting the evaluation as being influential (Herbert, 2011). Being an outsider to both the organisations, my research does not suffer from this limitation. Secondly, I noted that *evaluation influence* research suffers from a lack of transparent method, particularly in conducting case studies (Herbert, 2011). Between my methodology section and the case study protocol (Appendix A), I have provided the reader with a clear sense of how I collected my data and why. While admittedly I was not able to carry out the protocol to the same extent in the St Marys evaluation as in the Brighter Futures evaluation, my method of data collection and analysis is clear, transparent, and replicatable. Thirdly, I have attempted to deal with another limitation of the literature – self-report – by including participants from across the organisation, and by using organisational documents as a point of comparison. While there is an incentive for participants to present their organisation as being evidence-based, under the conditions of anonymity many participants were critical of evidence-based practice as it was implemented by their organisation, and of the concept itself. The inclusion of a broad set of staff, including direct service staff, team leaders, and managers across different areas that the programs were delivered in meant that a broader set of perspectives on the evaluation were available. The inclusion of extensive organisational documents, including evaluation reports, was important in grounding the participants’ reports of how the evaluation had been influential. Finally, I identified the importance of attention to the
timeline of the research in relation to when the evaluation had been completed. Accordingly I sought out evaluations that were ongoing, or had been completed in the past year.

While the process of applying the *evaluation influence* framework to the impacts was fairly intuitive, it was striking that there were relatively few individual level impacts. This is in contrast with other studies (Diaz-Puente, et al., 2009; Diaz-Puente, et al., 2008; Fjellstrom, 2007; Weiss, et al., 2005) that highlight the expectation that change processes start from individuals, who then influence others, which then connects to collective level change. Indeed the focus of this study on practice level learning was premised on the idea that practitioners' processes of self-reflection and questioning (Antonacopoulou, 2001) is what leads to practice change rather than collective level directives and cultural change. What was found in this case was much more top-down, with change happening at the Community Services level, or at high levels of The Benevolent Society, which then fed back down to individuals who drew on them to discuss or implement evaluation findings. While these collective level processes are likely to have had an individual and interpersonal process before it, these were simply not apparent in the data.

The finding that there was limited individual level influence is consistent with the review of *evaluation influence* studies (Section 2.7); individual and interpersonal level mechanisms become much more obscure when the study of influence begins after the evaluation has concluded, an artefact of the research design. As such, much of what participants recalled was at the collective level, with limited connection to personal impacts. Even when participants were asked directly about the impact the evaluation had on their practice or their attitudes and values about the program, they referred to policy and guideline changes handed down from higher up in the
organisation. In this way the expectation that influence begins at the individual level (Henry & Mark, 2003; Mark & Henry, 2004) did not play out in this case, although given the time gap between the conclusion of the evaluation and the beginning of the case study, it is possible that these mechanisms were obscured. It should also be noted that despite setting out to interview groups of managers and practitioners about the evaluation, participants in the practitioner group were often in leadership roles that gave them more of an insight into the dissemination and implementation of the evaluation at the level of the organisation rather than the individual.

A significant argument for the shift to the study of *evaluation influence* is the ability to map the pathways of influence, in order to better understand the connection between the different events, and how evaluations influence change. Early in the planning of this research I anticipated being able to develop a map of the pathways of influence, such as in Oliver (2008, p. 124). As discussed in the review of *evaluation influence* studies (Section 2.6), some researchers found this exercise challenging (Chang, 2006; Weiss, et al., 2005), Weiss et al. (2005) in particular describe "getting bogged down in unique tangles of strings (of influence)" (p. 26). This was very much the experience of attempting to map how influence played out in this evaluation. While the policy level impacts were clear, the processes that led up to them were missing from the accounts of the participants. In part this reflected the fact that many of the most significant decisions about the program sat outside The Benevolent Society. Another challenge was the processes that presumably precede the mechanisms identified by participants. For example while it was discussed that internal policy level learning occurred, such as the workshops around decreasing early exits, individual practitioners did not identify this as a change. Indeed examples of individual level changes in practice were rare; participants were not able to
identify specific changes in practice that came about as a result of the evaluation.

One of the advantages of evaluation influence research conducted concurrently with an evaluation (e.g. Diaz-Puente, et al., 2009) is that it is possible to closely observe influence from the beginning of the evaluation, particularly some of the individual level changes that culminate in collective change.

Section 3.2 set out a definition of practitioner learning, arguing it to be an individual phenomena that may be influenced by organisational values and processes. The analysis of influence across both cases found some examples of individual change, but primarily collective change that fed back into practice.

Talking directly to practitioners, individual changes reflected a change in attitudes towards evaluation (in both cases most of the things that had happened made practitioners more open to working with evaluation), providing a better understanding of how the program operates, and confirming existing expectations of the efficacy of the program (Sections 5.2.2 & 5.3.2). However much of the collective level changes that impacted on practice in both cases were large scale operational issues, where the impact on day-to-day practice was fairly negligible. While the practitioners had autonomy to determine how they practiced to an extent, particularly being based in NGOs, the practitioners in the research often found their practice influenced by requirements from different parts of the organisation, or from outside organisations (e.g. Community Services). The degree to which these changes were resisted or implemented superficially by practitioners – i.e. Meyer and Rowan’s (1991) decoupling and the logic of confidence – was unclear. At this point it also becomes unclear to what degree the evaluation can be said to have influenced program change, as opposed to the myriad of other influences on policy and program management. While a focus on practitioner level influence is important to the
development of practice-oriented evaluation (Schwandt, 2005), it remains a complex and convoluted area to research.

### 6.6 Limitations of the Research

The aim of this research was to develop theory based on participants’ experiences of a phenomena. In developing this theory, a number of limitations need to be acknowledged. These go to the scope of the research, ways the research was undertaken, and things that occurred over the course of the research.

The research set out to explore practitioner learning in human service NGOs. While this was done, it should be acknowledged that the analysis involves participants only from the NGO, and not the evaluators, and not - in the case of Brighter Futures - the funder and co-provider of the service. The decision to limit the scope of the research to just the NGOs in this case was deliberate; dealing with conflicting accounts of what happened and why across up to three organisations involved in the evaluation would have proved complicated, and difficult to manage ethically. While certainly it would have been interesting to get the perspective of people in other organisations about what happened in these evaluations, this was somewhat out of the scope of looking at practitioner experiences and perceptions of evaluation for one thesis. While this is a limitation in terms of providing a complete perspective on what occurred, I do not believe it is a limitation in terms of understanding practitioner learning.

The research concerned practitioner learning, yet much of the influence identified by the application of the Mark and Henry (2004) framework suggested that change primarily happened at the collective level. While capturing the broad narrative of the evaluations and their influence well, what was missing was a more detailed sense of how the experience of and information from the evaluation
impacted on day-to-day practice. That said, a number of individual level impacts were found around willingness to engage with evaluation and a better understanding of the program. The narrative of the evaluation is important to understand, but a more in-depth understanding of individual influence among practitioners may have been possible through more detailed interviews, or with interviews coinciding with evaluation dissemination or workshop activities.

Of the two case studies completed, one of them comprised of less participants and did not include the period following the release of the final evaluation report. As highlighted earlier in the thesis, this meant that a different experience of the evaluation had been captured, one that emphasised the participants’ experiences of the evaluation being conducted rather than its dissemination and implementation. While this limits the presentation of the full story of the evaluation and its eventual full influence, the inclusion of this case study was important to understanding practitioners’ perceptive about equitable and participatory engagement.

6.7 Chapter Summary

This chapter presents the key themes drawn from the case studies that describe the factors associated with practitioner learning from evaluation. The key theme seems to be the relationship between evaluation and practice; in summary, it is not clear to practitioners exactly how evaluation is supposed to inform their practice. For evaluation information to be considered by practitioners, it should have validity to inform practice, a criteria beyond a researcher’s understanding of validity and reliability. Practice validity refers to a clear connection and compatibility between the way an evaluation is undertaken, and the lived experience of practitioners. Evaluation should acknowledge the politics of evidence, the multiple and contested meanings of programs and evaluations, and the connection to broader issues about
evidence based practice. The identification and resolution of potential clashes of *inter-organisational priorities* are important, particularly in a service environment that removes practice knowledge from the funders of program and evaluations. These findings reflect broad themes of which many of the factors identified in the literature relate.

Drawing on these findings a model of evaluation to influence practitioner learning was proposed. The important components of this include, first embedding evaluation into practice, making evaluation information a tool for practitioners in order to help them understand and reflect on their effectiveness. Secondly, creating evidence that is valid to inform practice, ensuring that practice is evaluated with methodological rigour, but also with a clear connection to practice. Thirdly, making effective use of practitioners as evaluation stakeholders, valuing experiential and practice knowledge, and treating practitioners ethically and equitably. Fourthly, engaging with the critical politics of organisations and the sector, ensuring that evaluations question implicit and taken for granted assumptions, and avoid marginalising stakeholder groups. Finally, advocating for practice oriented evaluation at the planning stages of an evaluation, evaluators should make a clear case for the value and benefit of this type of evaluation.
7. CHAPTER 7: CONCLUSION

7.1 Introduction

This thesis set out to explore the conditions and barriers to practitioner learning from evaluation, drawing on the experiences of practitioners in evaluations of human service programs delivered by NGOs. Recognising a gap in the existing literature in exploring evaluation influence in the human service practice context, the research aimed to develop some guidelines for undertaking an evaluation based on the experiences and perspectives of practitioners themselves. What was found was fairly consistent with the factors identified in the evaluation use literature (Section 3.4), research that primarily reflects evaluation use for decision-making. Overall it suggests a desire to see evaluation collaborate with human service practitioners, to build their capacity to understand and participate in evaluation, and to create findings with practice relevance. Practitioners were critical of evaluation that fails to reflect the context of their work, that provides no direct benefits for caseworkers or clients, that is based on a picture of a program that reflects a particular interest group, and that is conducted in a way that suits external rather than internal interests and development.

The research ties together the idea of practitioner learning with evaluation influence, favouring influence over evaluation use. Evaluation use is subsumed within the influence mechanisms described by Henry and Mark (2003), and Mark and Henry (2004), represented by mechanisms in the behavioural level of analysis (e.g. policy change, program continuation, cessation or change). Importantly evaluation influence is concerned with the processes of change that lead to and follow on from changes in behaviour as a result of the evaluation, and emphasises the
importance of individual actors in the chains of events that lead to program change. Practitioner learning fits into evaluation influence as a way of focusing on the processes of thinking, reflection, attitude change, and behaviour change particular to individuals that work directly with vulnerable groups. This thesis argues that because of the autonomy and discretion available to practitioners to deal with the chaotic and personal nature of human service practice, influencing practice should be considered as primarily an individual process. That is, while evaluation can influence policy and practice guidelines, evaluation must have credibility and relevance to human service practice in order for these changes to connect to changes in human service practice. The findings of the research seemed to highlight this point, that the practitioners had significant concerns about the way the evaluations had been conducted that limited their credibility in informing practice change, despite the significant influence both evaluations had at an organisational level and a cross-organisational level.

For evaluation to achieve credibility and relevance in the human service context there is a need to recognise the broader politics and contestation in human service practice. The changing relationship between the government funders of human services and the NGOs that deliver these services is a cause for concern among practitioner groups. In particular, concerns about NGOs becoming ‘little bureaucracies’ due to their need to seek legitimacy from government (Balderstone, 2008), and the loss of focus on social justice and social democratic rights (Dolgoff, Loewenberg, & Harrington, 2009) present as particular concerns in the literature. Consistent with this, participants talked about evaluation as something that had to be done in order to be accountable to the funders of services, which lacked connection to the practice context. Particularly in the politicised and contentious space of human service practice, evaluations cannot hope to remain value neutral or objective, as
suggested in the findings there is a need to question implicit assumptions about the
program and the evaluation, and to ensure that different stakeholder perspectives are
acknowledged in the evaluation. Considering the importance of individual processes
of learning in human service practice in this research, there is a need for evaluations
to demonstrate a commitment to learning in order to challenge the sense of cynicism
that can exist about evaluation. This seems to be particularly important in the NGO
human service context where cynicism about evaluation and the interests it serves
may be exacerbated due to the sometimes-challenging relationships between NGOs
and the government funders of human services.

The thesis began with a discussion of the broader trend for human services to
be contracted out to NGOs (Section 1.7.1), a trend that comes with increasing
evaluation and a changing relationship between the funders and providers of
services. While practitioners often link these trends to the increasing managerialism
of practice (Carman & Fredricks, 2008; Taut & Alkin, 2003), they do have a keen
interest in their own efficacy and engage in self-evaluative processes (Cortis, 2006;
evaluation approach that values human outcomes and enhances the ability of
practitioners to deliberate well. This is argued as the best way to create change that
improves service delivery, the experiences of service users, and by corollary,
outcomes for service users.

Evaluation as a trans-discipline (Scriven, 2003) has been responsive to the
need to create information that can be used to improve programs and service
delivery. Participatory processes and the building of evaluative knowledge and
capacity among human service practitioners have become mainstream in evaluation
(Ebrahim, 2005a; Hoole & Patterson, 2008; Torres & Preskill, 2001). Section 2.4
describes how this is reflected in evaluation research, and the gradual shift to a broader definition than evaluation use. Attention to how evaluation affects the broader context of programs has in part led to the development of evaluation influence, an approach to understanding the broad, obscure, and diffuse impact evaluation can have (Henry & Mark, 2003; Mark & Henry, 2004).

Much of the evaluation use and influence literature has concerned the impact of evaluation on decision-making, this thesis argues that human service practice represents a different setting (Section 3.2). Practice is responsive, personal, and instantaneous, seldom allowing time to consult research or evaluation evidence. While organisations can set the values, and can create systems to legitimise and disseminate information (Pawlowsky, 2001; Schulz, 2001), ultimately the individual practitioner has the discretion to ignore, accept, or comply with organisational directives (Antonacopoulou, 2001; Brignall & Modell, 2000; Chang, 2006). This thesis suggests that evaluation influences practitioner learning, an individual process of reconciling evaluation with lived experiences and values (Antonacopoulou, 2001, 2006). This practitioner learning can then result in deliberate practice change.

While primarily reflecting the decision-making context, and evaluation use rather than influence, there is extensive research on the types of factors that affect the use/influence of evaluation. These are explored in Section 3.3. Drawing on Cousins and Leithwood (1986), Johnson et al. (2009), and other research (Alaimo & Van Slyke, 2006; Alexander, 2003; Avery & Van Tassel-Baska, 2001; Balthasar, 2009; Botcheva, et al., 2002; Burr, 2009; Christie, 2007; Diaz-Puente, et al., 2009; Diaz-Puente, et al., 2008; Fleischer & Christie, 2009; Grob, 2003; Karan, 2009; Ledermann, 2012; Murphy, 2007; Oliver, 2008; Preskill, et al., 2003; Sridharan & Lopez, 2004; Taut & Alkin, 2003), a set of factors were identified in the categories
of evaluation implementation, stakeholder involvement, and program or organisational setting.

Drawing on an abductive research strategy (Blaikie, 2007; Given, 2008), the thesis involved a three stage process: the presentation of case studies of evaluations that aimed to foster practice learning (Sections 5.2.1-5.2.3 & 5.3.1-5.3.3); an analysis of how evaluation influence played out in these cases, highlighted by the application of Mark and Henry’s (2004) framework (Sections 5.2.2 & 5.3.2); and the practitioners’ responses to the evaluation (Sections 5.2.3 & 5.3.3) that form the basis for the analysis of the conditions and barriers to practitioner learning (Section 6.2). Sections 5.2.1-5.2.3 concerned the evaluation of the Brighter Futures program, which was being evaluated on a large scale across sites and organisations. The Benevolent Society experienced considerable obstacles in getting access to data to inform practice once it became clear that they would have to take responsibility for fostering practice improvement themselves. Ultimately the experience was formative in the way evaluation was used and conducted in the organisation. The Burnside case study (Sections 5.3.1-5.3.3) concerned the evaluation of an integrated service model, the St Marys Child and Family Network. This research concluded before the final dissemination of the evaluation report, resulting in the case study being more about the experience of the process of the evaluation. While practitioners felt engagement had been done well up to a point, the level of consultation fell short of their expectations after a number of critical incidents that undermined engagement.

Drawing on both the organisational documents and the accounts of the participants, the influence of each of the evaluations was examined using Mark and Henry’s (2004) framework of evaluation influence mechanisms. This framework describes change processes at the individual, interpersonal and collective level. The
analysis of *evaluation influence* found primarily collective level impacts in each of the evaluations, with some broader individual level impacts around practitioners’ understanding of the structure of the program and their willingness to engage with evaluation.

In the analysis of *influence* in the Brighter Futures case study, some of the major changes were: the development of a set of norms and expectations that both practitioners and managers would use evaluation in their work; increased awareness of some of the different contexts the program operated in; and the development of some of the organisational processes around managing evaluation. The evaluation also seemed to have influenced a number of significant changes at the Community Services level, which led to the government department side of Brighter Futures being wound down. In The St Marys Child and Family Network evaluation some of the major changes included: the expansion of the integrated service model approach across other centres run by Uniting-Care Burnside; and a re-thinking of the program model based on the limited developmental change found by the evaluation.

Participants shared their perceptions and experiences of the evaluation, including their thoughts on the reasons for the limited influence on practice; these were drawn on in the analysis of the conditions and barriers for practitioner learning from evaluation.

### 7.2 Key Findings

The case studies provided a context to interview participants about evaluation and its relevance to human service practice. While the individual level impacts may have been limited, analysis of the cases and the participants’ perceptions and experiences of these evaluations suggest some key themes to practice oriented evaluation. These themes had a relationship to the factors in the literature associated with *evaluation*
use/influence, however these were more descriptive and less explicit that the factors in the literature.

7.2.1 The relationship between evaluation and practice
Underlying the findings was the issue of a lack of clarity about the relationship between evaluation practice, how exactly evaluation information is supposed to be incorporated into practice. While organisations have clear commitments to evidence based practice, and procedures and processes in place to promote it, there is a gap between the rhetoric and how an evaluation might actually translate to practice change. Practitioners thought that evaluations often ask questions not relevant to practice, produce findings not surprising to individuals who work within the framework of the program, and do not acknowledge their experiential knowledge of how programs function. The focus of evaluations tend to be on policy and organisational settings, and lack the meaning and context of program delivery. Practice takes place in a chaotic context, and is personal and intuitive, there is limited opportunity to consult evidence in the way a decision-maker or policy-maker might. In this way developing practice might be better thought about in terms of developing the capacities and competencies of practitioners.

7.2.2 The validity of evidence to inform practice
For evaluation information to be thought about, internalised and adopted into practice, it needs to be construed as valid by the practitioner (Fleming, 2011). Yet this includes more than validity and reliability in the sense that researchers and evaluators are familiar with; it requires a connection to practice. Practitioners understand the program frameworks they work within very well, to tell them something they do not know, or something that contradicts their practice experience requires evidence with a high level of credibility. This requires a mix of evaluation quality, credibility and relevance. The practitioners in the case studies provided an
extensive critique of the evaluations, primarily highlighting the failure of the evaluation to reflect the program, and the client outcomes that were apparent from the practice context. This impacts on stakeholder engagement in that where practitioners do not feel the evaluation reflects the practice context they are unlikely to engage, which may minimise influence. Working with practitioners and involving them with the design of an evaluation and the interpretation of results may enhance practice validity.

7.2.3 Equitable and participatory engagement
The participants suggested that evaluations need to draw on practitioners as valued stakeholders of an evaluation in order for to foster influence, when they are not engaged evaluation may engender indifference or even resistance. Valuing program and implementation knowledge can enhance the practice validity of an evaluation, but can also develop the long-term capacity for practitioners to understand and employ evaluative thinking in their practice. This engagement needs to be planned into the evaluation, and be reflected in the actions – not just the rhetoric – of evaluations.

7.2.4 Acknowledging the critical politics of evaluation
Evaluations are often associated with the requirements for accountability to funders that NGOs have to comply with, and thought of as part of the myth and ceremony of organisations (Meyer & Rowan, 1991). To engage practitioners, evaluations should acknowledge the critical politics of evaluation or else risk being considered a tokenistic or managerialist process that has no relevance to the real work of human service practitioners. This includes acknowledging critical issues about the evaluation, the program, and about broader issues of human service practice. Where the intent is there to foster learning and build evaluation capacity, this has to be clearly signalled to practitioners in order to distinguish the evaluation.
Acknowledging the perspective of practitioners and building critical questions into the evaluation may be a way of signalling this.

### 7.2.5 Resolving potential conflict between inter-organisational priorities

As the trend continues for government departments to purchase service outputs from NGOs (Australian Bureau of Statistics, 2002, 2009), these organisations become increasingly distant from the context of practice. Due to this distance, funders may become increasingly myopic about practice (Ebrahim, 2005a), seeing it as part of what the NGOs are responsible for. Despite practice improvement being in the interests of all parties, organisations responsible for the funding of services can be more concerned about accountability in an evaluation, and effectiveness in having their priorities represented in the evaluation. It may affect practitioner engagement if an organisation is unable to get practice improvement included as a priority in the evaluation. Also evaluations can be implemented poorly if there are an unclear set of roles and responsibilities for the different parties.

### 7.3 Recommendations

The analysis of the conditions and barriers to practitioner learning from evaluation suggests a number of clear recommendations.

### 7.3.1 Commitment to practice-oriented evaluation from funding bodies

Producing findings that have implications for practice, and developing the capacity of practitioners to work with evaluation information and to use evaluative thinking in their work is clearly in the interests of all parties. While undoubtedly this requires additional resources, additional time to conduct consultations and extra effort to engage practitioners, this kind of evaluation is cumulative; it builds long term capacities and competencies in individuals and organisations, and potentially leads to service delivery improvements. Committing to evaluation that is practice oriented is
a responsibility for the funders/commissioners of evaluations, and serves their own interests in improved service outcomes. While the delivery of the service has been contracted out, Government agencies still have a responsibility for quality service delivery which should be reflected in the evaluation of programs.

7.3.2 **Demonstrate organisational commitment to practitioner learning from evaluation**

While undoubtedly the funders of evaluations have a role to play in ensuring that evaluation has relevance to practitioners, the NGOs delivering programs also have an imperative to encourage and manage practitioner involvement in evaluation. The participants wanted evaluation to be embedded in practice, to have data enhance reflective processes, and to have evaluation capacity shared across the evaluation. The Brighter Futures evaluation was formative for The Benevolent Society, because of the way the evaluation was carried out an evaluation policy was developed and implemented, and the PIP process was piloted in Brighter Futures and rolled out across the organisation. These broad changes reflect efforts to encourage evaluation that has an impact at the practice level. The next level seems to be developing and encouraging evaluation capacity among practitioners, encouraging a culture of drawing on data in supervision and reflective practice. However, this requires ongoing engagement to ensure data reflects the practice context, and that there is capacity within programs to respond to data.

7.3.3 **Conduct evaluation in a way that enhances practice influence**

A key argument around social betterment (Mark, 2000) is the idea that like social programs, evaluations should be engaged in attempting to create a better social situation. While evaluators are inevitably contracted to undertake their work in a particular way, they also have significant influence on evaluation design. Based on the findings of this research, to improve practice influence evaluators should:
• Promote the embedding of evaluation into practice: While resourcing and creating systems to embed evaluation into practice are the responsibility of organisations, evaluators play a role in pointing out the benefits of this and suggesting ways to engage practitioners in evaluations they are contracted to undertake;

• Create evidence that is valid to inform practice: Ensure that practice is evaluated in a way that is methodologically valid, but also retains clear links to the practice context;

• Make proper use of practitioners as evaluation stakeholders: Properly valuing practitioners and stakeholders, incorporating their program and implementation knowledge, and treating them ethically and equitably;

• Engage with the critical politics of organisations and the sector: Encourage the questioning of implicit and taken for granted assumptions, and avoid marginalising some stakeholders;

• Advocate for practice relevant evaluation: Communicate for the mutual interest in producing findings oriented to practice level learning.

7.3.4 Continuing research into how to overcome the challenges of integrating research and evaluation in practice

This research was particularly novel as it addresses the challenges of evidence based practice in the context of evaluation research. In bridging the gap between evaluation and practice (Nutley, et al., 2003, 2007), and while it is important to conduct research and evaluation that is rigorous (Productivity Commission, 2009), there is also an imperative to create evidence that is useful. I would advocate for the importance of both large scale controlled trial evaluations (particularly when they are preceded with the kinds of measures outlined by Trochim (2009b), and evaluations that are focused on the practice context. That said there is a clear need for continued research into how evaluation can engage with the practice context.
7.4 Future Research

As suggested in section 7.3.4, there is a need to continue research about how evaluators can effectively engage practitioners and produce information that has value in thinking about human service practice. With this research having developed a sense of how practice oriented evaluation could be conducted, there is a need to further refine and develop this model, leading into some large scale empirical observations of evaluations, and simulation studies such as that conducted by Christie (2007). Collaborative research involving practitioners developing components of a model of practice oriented evaluation may also be valuable. There is also a need to further explore individual learning from evaluation as a psychological process, something that has been enthusiastically taken up by Mark, Donaldson, and Campbell (2011). Evaluation influence remains a conceptual framework that shows promise, but which requires continued research to ensure it remains viable as a way to think about the impact of an evaluation. There is a need to consolidate existing conceptual understandings and research approaches of evaluation influence towards some consistent definitions and procedures. This will help to develop the use of evaluation influence as a way of researching the impact of an evaluation. Further research on understanding the connection between different mechanisms of influence (i.e. the pathways of influence) may also help develop evaluation practices that are effective in creating change through specific mechanisms.

7.5 Conclusion

The central question in much of the discussion of research and evaluation utilisation (Nutley, et al., 2007) is about push/pull factors; if the difficulty in bridging research and practice is due to the inability of researchers to communicate findings, or the reluctance and inability of practitioners to accept research findings. This debate has a
very different dimension in the neo-liberal/managerialist world where research and evaluation knowledge are privileged sources of legitimacy (Goerke, 2003). NGOs in particular, given their position as the providers of social services, are subject to evaluation for accountability purposes, or engage in evaluation to seek legitimacy from potential funders. As Ebrahim (2008a) argues, evaluation can be part of a myopic view to accountability in the sector, and sometimes also disruptive and invasive; akin to pulling seedlings out of the ground in order to examine how they’ve grown. Yet despite all this, the participants in this research were enthusiastic about the potential for evaluation to help them reflect on their practice. While many were positively disposed simply because the evaluation was thought to improve the prospects of the program, many others also expressed an interest in understanding their effectiveness as practitioners, and how they might work better in order to get better outcomes for their clients.

While at the beginning of this research I was drawn to the simple dichotomy of push/pull factors, framing the discussion in this way misses the subtlety of what actually happens when programs are evaluated. Evaluation cannot be value neutral or objective, it is commissioned for someone and for some purpose; inevitably those who commission and act as stakeholders of an evaluation, with some purpose in mind. While they may engage other stakeholders, and undertake consultation in order to expand the legitimacy of the evaluation and to foster its use, the core purpose of the evaluation being commissioned is unlikely to be compromised.

Evaluating services in a way that is useful for practitioners is a win-win. It can foster improvements that have meaning and impact to service users, a culture of excellence in practice, and the potential to empower human service practitioners in their work. The challenge is in confronting the tendency to accountability myopia
(Ebrahim, 2008a) by the commissioners of evaluations. The danger lies in organisations like Community Services becoming increasing removed from direct practice, and merely funders or managers of human service provision (O’Shea, 2007; O’Shea, Leonard, & Darcy, 2007). As program implementation and practice knowledge is lost from the organisation, this myopia is likely to increase. The responsibility for practice improvement and service quality is likely to increasingly fall to the deliverers of human services, NGOs like those included in this research. While each of these organisations had a strong research and evaluation capacity, and a history of conducting formative and practice oriented evaluations, scrutiny of NGO spending on governance and administration may challenge this capacity.

My conclusion is that the funders of human services should have a direct interest in supporting practice-oriented evaluation, to ensure continuous improvement and increased value for money from the service. The model developed from this research (Section 6.4) provides one such approach to undertaking this type of evaluation. While this work can be challenging for evaluators, collaborating and achieving consensus or compromise is fundamental to contemporary evaluation (Owen, 2007). NGOs and evaluators should be strong advocates for evaluation that has meaning and significance for practice, and challenge a short-sighted view of evaluation for accountability. The potential for social improvement in the lives of vulnerable service users argues for the application of evaluation to the challenges of human service delivery.
REFERENCES


Allen, M. (2010). Steeping the organisation's tea: Examining the relationship between evaluation use, organisational context, and evaluator characteristics. Doctor of Philosophy, Case Western Reserve University, Cleveland.


Case Study Protocol

1. Background

Approaches that foster learning and program improvement feature prominently in contemporary evaluation theory, primarily due to the potential of such approaches to improve social conditions for service users (Schwandt, 2005; Shaw & Shaw, 1997; Taut, 2007). Despite the enthusiasm for this more formative, participatory, and pedagogical approach to evaluation, there are considerable difficulties in fostering learning from evaluation in Non-Government Organizations (NGOs) that provide services funded (in part) by state and federal governments. Primarily, evaluations of these services tend to be oriented to providing upwards accountability (Ebrahim, 2005), and the types of information important to funders (Carman, 2007; Hoole & Patterson, 2008). Even with an explicit aim to foster learning, evaluations can be confounded by the complex institutional context that exists within NGOs, situated in a wider milieu of contestation about the role of NGOs as service providers and their relationships with service funders (Maddison, Deniss, & Hamilton, 2004; O’Shea, 2008). Added to this, the traditional unfamiliarity of evaluation procedures in the sector (Mcdonald, 1999; Taut & Alkin, 2003), and a strong preference for tacit and experiential learning amongst human services practitioners (Healy & Meagher, 2001), there are some clear cultural and institutional barriers to building evaluative capacity within NGOs.

The proposed research will examine the personal and organisational experiences within NGOs that have undergone an evaluation that at least in part aimed to foster internal learning, in order to develop understanding about the conditions and processes that may help or hinder learning. This will be pursued by achieving an understanding of: (a) the context of the evaluation, including the nature of the program and organisation, and how the evaluation played out in this institutional environment, which will in turn provide a context for the; (b) lived experiences of the people involved in and affected by the evaluation.

2. Aims and Outcomes

The broad aim of the research is to inform the use of evaluation as a means of fostering learning and improvement in social programs run by NGOs. More specifically the research aims to:
APPENDIX A

- Obtain a comprehensive understanding of two evaluations, which was pursued through a simple case study (Stake, 2003);
- Use Mark and Henry’s (2004) evaluation influence framework to identify instances where the evaluation has affected change; and
- From the evaluation influence mechanisms identified, conduct a phenomenological analysis of the experiences of participants in order to infer a set of conditions and barriers for practitioner learning from evaluation, and consider the implication of these for evaluation practice.

The primary outcome of the research is the production of a thesis report, which will serve as the source for several peer reviewed journal articles and conference presentations, and case study briefs for each of the participating organizations.

3. Methodology
The research will take place across two New South Wales NGOs, specifically targeting human service based programs that:

- Are sustained and established programs that have reasonably secure funding and ongoing support for their continuation;
- Are partly or primarily funded by an external agency to which the NGO is accountable to;
- Involve a human interaction between staff and clients in a community setting as the central program activity; and
- Have been evaluated within the past year (or have ongoing efforts to promote the use of previous evaluations) and the evaluation at least in part identified internal learning and improvement as a goal of the evaluation.

a) Research Methodology:
The research aims will be pursued through two case studies, which will then be analysed in three stages to fulfil the research aims. The data will be collected in a single stage, with documents and interviews obtained over the course of a year.

A case study is an empirical enquiry that investigates a phenomenon within its real life context, and is especially useful when the boundaries between the phenomenon and the context are not clear (Yin, 2009, p. 13). My intention is to use the case study approach as a meta-evaluation; a study of the evaluation study; a purpose for which a
APPENDIX A

case study is well-suited (Green, 1992; Luck, Jackson, & Usher, 2006; Yin, 2009). The case studies serve as the narrative of the evaluation; from the context that existed prior, the impetus for the evaluation, the activities and events that constitute it, to everything that occurred following. The work of Mark and Henry (2004), and Henry and Mark (2003) in developing the concept of evaluation influence provide a framework for data collection of the case studies as well as providing some theoretical perspectives (mechanisms of evaluation influence) to pursue in the analysis of the data. The case studies will provide the context with which to understand the experiences and perceptions of the service level staff affected by the evaluation.

Interviews with the service level staff will serve to make their perspectives on their work, the organisation, evaluation, and learning through evaluation explicit. These interviews will be phenomenological in the sense that the focus of the interviews is on "...the structure and essence of experience of this phenomenon..." (Patton, 1990, p. 69), the phenomenon being the evaluation of their program/service. Consistent with the epistemological and methodological approach of the research, a hermeneutic/interpretative approach will be employed in the collection and analysis of the interview data (Blaikie, 2007; Lopez & Willis, 2004; Racher & Robinson, 2003). This is an approach characterised by the goal to obtain a deep understanding from the participants’ everyday experiences (Broussard, 2006), and to identify the underlying structures within these experiences (Banaga, 2000). The aim of these interviews are to obtain a sense of the participants’ experiences and meaning, a sense of the life-world they inhabit, and an understanding of the relationship between these social, cultural, historical, and political contexts.

b) Research Method within each Organisation:
The proposed research will take place within two different organizations following a similar research process. Once permission to undertake the research has been obtained, the case study of the evaluation will be undertaken, with the phenomenological interviews with service level staff.

The case study will be a process of building an understanding of the organisation, the program, the evaluation, and the environment the services operate within. The process will begin with obtaining organizational documents (annual reports,
evaluation documents), notes from meetings and from observations, and leading into interviews with key stakeholders involved in the evaluation. It is anticipated that interviews with approximately five staff members who have key experiences of the evaluation and the program will be adequate (complemented by the organisational documents) to be able to form a clear descriptive account of the program and evaluation. These informal interviews will be relatively brief (20-30 minutes) and will directly concern the goals of the program, the narrative of the evaluation, and how evaluation influence played out in the organization.

The phenomenological interviews will involve 5-10 service level staff (staff who are involved in directly engaging with service users) in a semi-structured interview of between 30-45 minutes in length. While the interviews undertaken as part of the case study sought to develop descriptive detail about the program and the evaluation, these interviews concern the participant’s experiences and perceptions of: work in the organization; the practice of human service work; evaluation in general; evaluation intended to foster internal learning; and of the evaluation that occurred and its impact in the organization. Interviews will be undertaken at the participant’s convenience at a location deemed satisfactory to the participant in terms of privacy and confidentiality.
Participant Information Sheet (General)

An information sheet, which is tailored in format and language appropriate for the category of participant - adult, child, young adult, should be developed.

Note: if not all of the text in the row is visible please 'click your cursor' anywhere on the page to expand the row. To view guidance on what is required in each section 'hover your cursor' over the bold text. Further instructions are on the last page of this form.

Project Title: Evaluation for learning: Conditions for flourishing in non-profit human service programs.

Who is carrying out the study?
James Herbert, a PhD student at the University of Western Sydney, supervised by Associate Professor Natalie Bolzan and Dr. Michael Houlbrook.

You are invited to participate in a study conducted by James Herbert, PhD student at the Social Justice Social Change Research Centre at the University of Western Sydney. The research will form the basis for the degree of Doctorate of Philosophy at University of Western Sydney under the supervision of Associate Professor Natalie Bolzan and Dr. Michael Houlbrook.

What is the study about?
The purpose is to investigate the experiences and impact of evaluation on the work of non-profit human service providers. The interviewer will be interested in how you feel about your work, what you thought of the evaluation process, what you thought about the results, as well as if the evaluation has changed anything about your work.

What does the study involve?
The study will involve a semi-structured interview, which will be recorded and transcribed. This transcription will be available for you to review at your request. The interview will be held at your convenience at a location that allows for your privacy and confidentiality to be maintained.

How much time will the study take?
Interview lengths will vary, but will most likely be between 30-60 minutes.

Will the study benefit me?
There are no direct personal benefits besides the chance to reflect on the way your organisation interacts with evaluation and to suggest how evaluation could be most useful in enhancing the way the organisation undertakes its work. There is also an indirect benefit in promoting the work of the organisation and the program.

Will the study involve any discomfort for me?
No, however suitable workplace counselling or grievance services can be identified if necessary. In most cases services will be available through your workplace or union (if you are a member), otherwise Anglicare, Centrecare, and the Salvation Army provide free telephone and face to face counselling.
APPENDIX B

How is this study being paid for?
The study is being sponsored by University of Western Sydney post-graduate research funding.

Will anyone else know the results? How will the results be disseminated?
All aspects of the study, including results, will be confidential and only the researchers will have access to information on participants. The research will be written up into a chapter of the student’s final thesis as well as considered for publication in peer reviewed academic journals. A summary report will also be offered to the organisation for their information. Participants will not be identifiable in any of these publications.

Can I withdraw from the study?
Participation is entirely voluntary: you are not obliged to be involved and - if you do participate - you can withdraw at any time without giving any reason and without any consequences. Alternatively you may choose to omit parts of any information you provided on reviewing the transcript of the interview.

Can I tell other people about the study?
Yes, you can tell other people about the study by providing them with the chief investigator’s contact details. They can contact the chief investigator to discuss their participation in the research project and obtain an information sheet.

What if I require further information?
When you have read this information, the researcher James Herbert will discuss it with you further and answer any questions you may have. If you would like to know more at any stage, please feel free to contact James Herbert (PhD Student): 0402296734 or 16459830@student.uws.edu.au

What if I have a complaint?
This study has been approved by the University of Western Sydney Human Research Ethics Committee. The Approval number is H8608.

If you have any complaints or reservations about the ethical conduct of this research, you may contact the Ethics Committee through the Office of Research Services on Tel 02-4736 0883 Fax 02-4736 0013 or email humanethics@uws.edu.au.

Any issues you raise will be treated in confidence and investigated fully, and you will be informed of the outcome.

If you agree to participate in this study, you may be asked to sign the Participant Consent Form.
APPENDIX C

Interview Schedule

1. Could you tell me a bit about your role and the connection between your work and the program?
   a. What does your day-to-day work involve?
   b. What is the work environment like?
   c. How does the program fit in with the work of the organisation?

2. The organisation has a stated commitment to evidence based practice, how does this play out at a practice level?
   a. Is research and evaluation discussed in supervision and/or practitioner groups?

3. Can you describe the program and what it is aiming to do?
   a. What are the important outcomes?
   b. How does the program pursue these in terms of human service practice?
   c. <Brighter Futures only> Is there any differences between your site and the other Brighter Futures sites?

4. <Brighter Futures only> How does the relationship with DOCS as the funder of the program and co-provider of the service work?

5. How did the evaluation play out, what was the narrative of the evaluation?
   a. Did you have any prior experiences of evaluation, did you know what the evaluation was for?
   b. What did you know about what was going to happen before the evaluation began?
   c. From start to finish can you tell me what happened, what parts of the evaluation were you involved in, what parts were you not?

6. What did the evaluation find and what did you think about these conclusions?
   a. How have the findings been disseminated to you?
   b. What has occurred so far in terms of working with and discussing the findings?

7. Overall, what has been the impact from the evaluation, what has occurred as a result?
   a. Has it changed your views of the program or its function?
   b. Has it changed anything about the way you do your work?
APPENDIX C

c. Has it featured or been used in discussions about the program?
d. Has it changed anything tangible in terms of policy or guidelines for practice?

8. For an evaluation to be most useful for practitioners, how would it be done, and what kind of information would it produce?
Participant Consent Form

This is a project specific consent form. It restricts the use of the data collected to the named project by the named investigators.

Note: If not all of the text in the row is visible please 'click your cursor' anywhere on the page to expand the row. To view guidance on what is required in each section 'hover your cursor' over the bold text.

Project Title: Evaluation for learning: Conditions for flourishing in non-profit human service programs.

I, .................................., consent to participate in the research project titled Evaluation for learning: Conditions for flourishing in non-profit human service programs.

I acknowledge that:

I have read the participant information sheet and have been given the opportunity to discuss the information and my involvement in the project with the researcher/s.

The procedures required for the project and the time involved have been explained to me, and any questions I have about the project have been answered to my satisfaction.

I consent to the interview, which will cover topics related to my working life and the impact of the evaluation in my work and within my organisation. I also consent to the audio recording of this interview.

I understand that my involvement is confidential and that the information gained during the study may be published but no information about me will be used in any way that reveals my identity.

I understand that I can withdraw from the study at any time, without affecting my relationship with the researcher/s now or in the future. I also understand that I am able to review the interview transcript prior to the publication of the research.

Signed: ____________________________

Name: ____________________________

Date: ____________________________

Return Address: ____________________________
APPENDIX E

Analysis Process

Case Study Analysis

- Review and analysis of documents describing the evaluation;
- Categorisation into descriptive detail on: (a) the organisation; (b) the program; (c) the evaluation;
- Reanalysis of the case study information drawing on interview data to provide alternative perspectives to the documented sources;
- Written presentation of the case to refer to over later stages of the analysis.

Identification of Mechanisms of Evaluation Influence

- Development of a set of definitions of the evaluation influence mechanisms (Section 4.9.2);
- Collection of all identified impacts of the evaluation from the interviews and organisational documents;
- Coding of influence events according to definitions;
- Review of coding including identification of any new categories and examples of impact that didn't fit into an existing mechanism;
- Presentation of the analysis of influence for each case;
- Analysis of the connection between influence events.

Phenomenological Analysis of the Experience of Evaluation

- Line by line analysis of interview transcripts;
- Development of provisional themes and coding of each case individually into: (a) Program context; (b) culture of evaluation in the organisation; (c) experiences of the evaluation (quality quality & credibility; perceived relevance; invasiveness; management of process; dissemination; good practice);
- Drawing on the case studies and the organisational documents to provide context for the participant accounts;
- Drawing on the analysis of influence to connect the experience of the evaluation with how influence played out in each of the cases;
- Development of macro level themes that describe the conditions and barriers to practitioner learning across both cases, and a model of good evaluation practice.