‘Do the data in fact deceive?’

An Analysis of the Roles of Evaluation and the Production of Aid Effectiveness at the World Bank



Alexandra Clare Yannias
School of Geography and the Environment
Wolfson College, University of Oxford
In partial fulfilment of the requirements for the degree of
Doctor of Philosophy
Hilary Term, 2015

Abstract

‘Do the data in fact deceive?’:

An Analysis of the Roles of Evaluation and the Production of Aid Effectiveness at the World Bank

This is a dissertation about the organisational structure of the World Bank, the professional practice of evaluation, and the meaning of the concept of ‘aid effectiveness’ in practice. In international development, evaluation is a professional activity that determines and then reports on the impacts of aid projects and programmes to the clients of such efforts and to the public. ‘Aid effectiveness’ is a concept that refers to a standard of how aid projects and organisations should operate and the results, such as economic growth and poverty alleviation, which these efforts should deliver in order to ‘work.’ The concept of ‘aid effectiveness’ has also been used in the debate about international development as a system and its reform. Given that aid policymakers and academic researchers often use the data contained in development organisations’ evaluations to determine the extent to which aid projects and programmes are ‘effective,’ it is critical to analyse what these evaluations measure and what influences their ratings and judgments.

Based on a case study of the World Bank, the analysis is primarily qualitative and draws on both interviews with evaluation professionals in the World Bank and content analyses of the logical framework, indicators, and language in the World Bank’s evaluations at the project- and country-level. Building on the previous theoretical work in post-structuralism that considers how international development organisations ‘produce’ their work through certain terms and processes (Escobar, 1995; Crush, 1995), I assess how the professional practice of evaluation in the Bank ‘produces’ the results of aid at the project- and country-level, specifically in the evaluation reports that it makes

publically available. The World Bank's data and evaluation reports are a 'window' through which to understand the impact of aid, and several factors that influence this 'window' are assessed, including the institutional role of evaluation, the professional practice of evaluation, and the required evaluation processes within the World Bank.

The study has important implications for practitioners of international development, academic researchers, and evaluation professionals who endeavour to improve the aid system and often rely on the results of the World Bank's evaluations to inform their understanding of the impact of particular development efforts. By reshaping the discussion from one which considers if aid 'works' to one about the data and the process of making a judgment about the success of aid projects and programmes, I articulate what the role of evaluation is in practice and what the World Bank's resulting evaluative data do and do not reflect about the World Bank's work. The relationship between the 'scales' of aid is also analysed by comparing and contrasting the evaluation processes at the project-level and the country-level. I challenge the notion of a 'micro-macro paradox' (Mosley, 1986) between the successful results of the World Bank's projects and the economic development in its client countries by articulating the actual meaning of this data in context, the unseen institutional forces that shape this data, and the difficulty of asserting a linear relationship between the results of projects and programmes on different scales of aid.

Table of contents

Abstract..... 2
Table of contents..... 4
List of tables and figures..... 7
List of acronyms..... 8
List of appendices..... 9
Acknowledgements..... 10

Chapter 1
Introduction..... 12

1.1. The focus of the research..... 12
1.2. Motivations and research topic..... 13
1.3. Research objectives and questions……..... 16
1.4. Summary of chapter contents…..…..... 19

Chapter 2
Reshaping the aid effectiveness debate:
A review of the relevant literature..... 27

2.1. Introduction..... 27
2.2. Development economics and the practice of development..... 28
2.2.1. Challenges within development economics..... 31
2.2.2. The Washington Consensus and the Bank today..... 33
2.3. Critiques of the results of international development..... 36
2.3.1. The micro-macro paradox..... 37
2.3.2. Other criticisms based on the results of aid..... 40
2.4. Critiques of the assumptions and processes of development..... 43
2.4.1. The narrative of development..... 44
2.4.2. Space, place, and scale..... 48
2.5. Institutional Analysis: Evaluation and evaluators..... 52
2.5.1. The role of the practice of evaluation within the Bank..... 53
2.5.2. The role and influence of professional evaluators 55
2.6. The theory of monitoring and evaluation..... 58
2.6.1. Dissecting the logic and language of the evaluations..... 59
2.6.2. Limitations of the Bank's evaluation process..... 62
2.7. Summary..... 64

Chapter 3
Analysing the 'window' of the World Bank's evaluation data: Methodology and theoretical considerations..... 67

3.1. Introduction..... 67
3.2. Qualitative methodology..... 67
3.3. Case study approach..... 69
3.4. Elite interviews..... 70
3.4.1. Selecting and conducting the interviews..... 71
3.4.2. Considering the interview environment..... 75
3.4.3. Interview questions..... 77
3.4.4. Analysing the interviews..... 79
3.4.5. Elite interviews and positionality..... 81
3.5. Content analysis and theoretical considerations..... 83
3.5.1. Selecting the documents for the sample..... 85
3.5.2. Analytical process for the content analysis..... 87

Chapter 4
Evaluation in context: Deconstructing the organisational role and spaces of evaluation at the World Bank..... 90

4.1. Introduction..... 90
4.2. Methodology for analysing the 'evaluation space'..... 92
4.3. Raison d'être: The evolution of evaluation at the Bank..... 94
4.4. IEG's independence and the evaluation space..... 99
4.4.1. The impact of IEG's independence on evaluation..... 101
4.5. IEG's regulative role..... 107
4.5.1. IEG's regulative role and the places of evaluation..... 111
4.5.2. Challenges to IEG's regulative role..... 117
4.6. Summary..... 124

Chapter 5
The art of evaluation: The evaluation process, evaluation professionals, and the production of aid effectiveness at the World Bank..... 126

5.1. Introduction..... 126
5.2. Methodology for articulating the 'insider's view'..... 127
5.3. The project-level evaluation process..... 130
5.3.1. Evaluators' explanations of project-level ratings..... 132
5.4. The country-level evaluation process..... 137
5.5. Elites in the production of aid effectiveness..... 141
5.5.1. Who?: Evaluators' views on themselves..... 142
5.5.2. How?: Evaluators' influences and experience..... 146
5.5.3. Why?: Evaluators on the goals of evaluation..... 150
5.6. Summary..... 152

Chapter 6
Codifying success in foreign aid projects: A content analysis of the World Bank's project-level evaluations..... 155

6.1. Introduction..... 155
6.2. Methodology and a review of the data..... 157
6.2.1. Sample selection..... 161
6.3. The meaning of the ratings: data and language..... 162
6.3.1. Development outcome and ERR..... 162
6.3.2. Sustainability of project outcomes and evaluation of actors.. 167
6.3.3. Sources of data used to assess the Bank's projects..... 171
6.4. The narrative of 'satisfactory' and 'unsatisfactory' ratings..... 175
6.4.1. Attribution in the logical framework..... 182
6.5. Summary..... 184

Chapter 7
The scale of success in international aid: Measuring success and defining attribution in the World Bank's country-level evaluations..... 186

7.1. Introduction..... 186
7.2. Methodology..... 188
7.2.1. Descriptions of the sources used..... 189
7.2.2. Defining the sample..... 192
7.3. The Bank's logical framework at the country-level..... 194
7.3.1. The Bank's logical framework in context..... 195
7.4. The Bank's goals at the country-level: What and how?..... 199
7.4.1. What?: The Bank's goals at the country-level..... 199
7.4.2. How?: Assessing the Bank's impact on poverty alleviation.. 203
7.4.3. From goals to country-level impact..... 207
7.5. Attribution in the Bank's country-level evaluations..... 213
7.6. Summary..... 218

Chapter 8
Data, deceptions, and conclusions……..... 220

8.1. Introduction..... 220
8.2. A conclusion in three parts..... 221
8.2.1. Data and the goal and role of evaluation..... 223
8.2.2. The evaluation process and the meaning of the ratings..... 226
8.2.3. 'Success' and attribution in context..... 231
8.3. Implications..... 233
8.3.1. Implications for the multi-disciplinary academic literature.. 234
8.3.2. Implications for development policy and evaluation systems 237
8.4. Directions for future research..... 239

Appendices..... 243

Bibliography..... 261

List of tables and figures

Figure 4.1.	The top of the World Bank's organisational chart, 2014	98
Table 6.1.	Rating explanation, World Bank's Harmonized Criteria	160
Table 6.2.	Data used for each rating	173
Figure 7.1.	Country-level evaluation process	189
Table 7.1.	Publically available World Bank country-level evaluations	193
Table 7.2.	Pillars in the evaluations included in the sample	201
Table 7.3.	Projects that evaluators use to assess poverty reduction goals	205
Table 7.4.	Top down indicators used in the Bank's evaluations	210
Table 7.5.	Bottom up indicators used in the Bank's evaluations	211

List of acronyms

AEL	Aid Effectiveness Literature
CAE	Country Assistance Evaluation
CAS	Country Assistance Strategy
CASCR	Country Assistance Strategy Completion Report
CPE	Country Program Evaluation
DAC	Development Assistance Committee
DIME	Development Impact Evaluation Initiative
ERR	Economic Rate of Return
ECG	Evaluation Cooperation Group
GDP	Gross Domestic Product
GNP	Gross National Product
ICR	Implementation Completion Report
IEG	Independent Evaluation Group
IFC	International Financial Corporation
IMF	International Monetary Fund
ISI	Important Substitution Industrialisation
MDG	Millennium Development Goal
MDB	Multilateral Development Bank
OED	Operations Evaluation Department
OECD	Organisation for Economic Cooperation and Development
PRSP	Poverty Reduction Strategy Paper
PDO	Project Development Objective
PPAR	Project Performance Assessment Report
SAP	Structural Adjustment Programme
TTL	Task Team Leader
UNEG	United Nations Evaluation Group

List of appendices

Appendix 1	Summary of Interviews	243
Appendix 2	Interview Questions	244
Appendix 3	Project-level evaluations (Implementation Completion Report Reviews) included in sample for content analysis	245
Appendix 4	All sections of the ICR Review, by information provided and rating	248
Appendix 5	Publically available country-level evaluations in African countries	250
Appendix 6	Publically available Country Assistance Strategy Completion Report (CASCR) in African countries	252
Appendix 7	Pillar Themes in Country Program Evaluation (CPE), Country Assistance Evaluation (CAE), and Country Assistance Strategy Completion Report (CASCR) Documents	253

Acknowledgements

At the end of my dissertation-writing, I find myself with a deep debt of gratitude to the many individuals who helped and motivated me during this process. Many of the positive attributes that one may find within these pages are due to those people named below and, of course, any errors that appear within these pages are my own.

I would like to begin by thanking my supervisors, Dr. Dariusz Wójcik and Dr. Linda McDowell, for their thoughtful supervision, for reading my drafts, and for providing me with countless helpful suggestions and research strategies. I have appreciated their thoughtful comments, encouragement, and seemingly endless supply of expert advice. At Oxford, I would also like to thank Dr. Gordon Clark and Dr. Craig Jeffrey for their helpful comments in my confirmation assessment and for helping me to improve the theoretical framework and the structure of the final work. In my viva, Dr. Clark and Dr. Adam Dixon (University of Bristol) provided thoughtful critiques in a thorough and challenging exam, for which I am also very grateful and which inspired interesting future research ideas. Finally, in the process of completing my research, I have relied upon Ms. Ruth Saxton, the department administrator, and Ms. Rachael Connolly and other administrators at Wolfson College. I am grateful for their thoroughness, patience, and responsive help throughout the years.

I would not have been able to complete this project without the help of the World Bank evaluators and the academic experts whom I interviewed. I am keenly aware of how fortunate I was that professionals with such demanding schedules replied to my interview requests and provided me with open and genuine feedback. I am thankful for their time and their willingness to assist me in my research. I must also thank the World

Bank administrative professionals and security personnel for their help in scheduling my interviews, applying for the evaluation conference, and in navigating the bureaucratic process of gaining access to the Bank.

The writing process is a lonely occupation and I am extraordinarily grateful to the many friends, colleagues, and relatives who have helped motivate and support me. I was able to complete my interviews in Washington, D.C. with the gracious hospitality of my friends Liz and Patrick, Melissa and Stacy, and Cheryl and Phil. My Oxford colleagues, Nihan, Eveliina, Helena, and Richard provided vital moral support in the difficult moments of research and dissertation writing as did the Wellesley Alum Community forum and my former advisor, Lidwien. I would also like to thank my very kind parents, my husband, and my friend Wendy for providing me with important revisions and helpful comments on my drafts. On a personal level, I am very thankful to my parents, Kate and Dean, and my brother, Nick, for their constant support and for helping me keep some of my sense of humour during this process and to my father-in-law, Gottfried, for his support of my studies. I am also grateful to Juliana, Gayle, and Libby for their inspiring grace and strength these past few years.

And last, and perhaps most, I am grateful beyond words to my incredible husband and friend, Andreas, for everything.

Chapter One

Introduction

1.1. The focus of the research

The World Bank ('the Bank') is an influential financial institution that has emerged since it was created in 1944 as not only a lead donor of international development projects, but also as a key source of information about the practice of international development (Grasso et al., 2003; Broad, 2006). This is a study about the World Bank's organisational practices, the professional practice of evaluation within the Bank, and the concept of 'aid effectiveness.' The analysis focuses on three issues: the professional practice of development evaluation in the World Bank, the ways in which the institutional role of evaluation influences the information that the Bank's evaluations are based upon, and the narrative that evaluations 'produce' about aid. By investigating what the Bank's evaluation processes measure and what data are considered in these determinations, I challenge some of the conclusions about the concept of 'aid effectiveness' that previous academic studies have put forth (Mosley, 1986). I argue that the Bank's evaluation processes often follow the models suggested by academic studies and international standards and, rather than being a straightforward reflection of if aid 'works,' the results are based upon a much wider range of institutional and procedural issues.

I investigate both the World Bank's practice of evaluation as well the narrative the Bank's evaluation teams 'produce' through their written materials and ratings, drawing on post-structural analyses of development (Escobar, 1995; Crush, 1995). In particular, I

explore how the institutional roles and professional evaluation practises in the Bank shape its evaluation results at different 'scales' of aid, which range in geographical focus from evaluations of the Bank's projects in particular communities to its country level programmes. The steps and processes by which the Bank's evaluation teams 'produce' evaluative data is a 'window' through which to understand the Bank's projects and programmes in the communities and countries in which it operates and these are often the data used to determine aid's impact more broadly. This dissertation is about this 'window' and the context in which evaluation data are created rather than the relative merits of any type of development programme or project. In this introductory chapter, the motivations, research questions, and research objectives are discussed. I then provide an overview of the chapters in this dissertation and explain how each chapter contributes to the overall research objectives.

1.2. Motivations and research topic

I was motivated to look in-depth at the roles and processes of aid evaluation by both my personal experience and by previous academic studies. During the process of writing my MPhil dissertation at the University of Cape Town on the right to housing in the South African Constitution, I worked as an intern for a team evaluating a self-help housing project. The organisation responsible for this project worked with the residents of an informal community to build formal structures within their existing community rather than forcing them to relocate to government-built housing (Huchzermeyer and Karam, 2006). I was inspired by a valid but difficult question posed to me while filling the role of an evaluator. One afternoon, while I was completing surveys in Mitchells Plain, a township outside Cape Town, a community leader asked me: "Why do you get to

decide if a project works here?" The emphasis in this question was on the words 'you' and 'here,' highlighting my position as a white, Western woman currently enrolled in a graduate programme at an elite university. Given my interest in evaluation and international development, this community leader's valid question motivated me to consider the role of the experts within evaluation and what function evaluation serves within development institutions, particularly powerful organisations such as the World Bank that are often the creators of development data.

The issue of 'positionality,' or the process of considering the position of the researcher in relation to the topic and to the research subjects is a key issue in the geographical literature and elsewhere (Rose, 1997; McDowell, 1998; Mullings, 1999), as is the question of the role of professionals within development institutions (Burawoy et al., 2000; Strathern, 2000; Broad, 2007). While there is significant previous research that discusses positionality and the role of professionals in organisations such as the Bank and the International Monetary Fund (IMF) (Strathern, 2000; Broad, 2007), there is less research on postionality as it relates to evaluation professionals as particular professionals within the international aid system.

I became interested in how the evaluation processes in which evaluation professionals engage, and their role as a unit within institutions such as the Bank, impact on the results of these evaluation reports. The initial questions I wanted to answer included the following: What are the professional practises involved in evaluation in international development organisations and how do the professionals who engage in this work explain and view their position? Where does the evaluation function fit in the large organisation that is the World Bank and what is the overall purpose of evaluation? As is discussed in the next section, the questions evolved as new issues emerged from the

literature, the interviews with academic experts and Bank professionals, and from the content analyses of the Bank's written evaluations.

The many debates about aid effectiveness and international development in economics, geography, and other social sciences also motivated me to pursue these questions (Mosley, 1986; Hyden, 2008; Essex, 2012). The previous research on aid effectiveness focuses primarily on either a quantitative analysis of the Bank's ratings or a theoretical analysis of the assumptions that inform the Bank's work. In this thesis, I shift the focus from the end result of the Bank's evaluative work (its evaluation ratings and the measurable impacts of its work on its recipient countries) to a consideration of what the data, upon which previous research is based, measure and how they are used. I directly engage with a key area of research on the topic of aid effectiveness which is referred to as the 'micro-macro paradox' (Mosley, 1986). Based on a quantitative analysis of the Bank's project-level ratings and economic growth at the country-level, Paul Mosley concluded that the Bank reported strong rates of return for most of its projects but that foreign aid, overall, had not led to economic growth in recipient countries (Mosley, 1986). In considering this relationship, Mosley (1986, 22) asked: "What is going on? Is it true as the data suggest that aid projects are succeeding while aid as a whole is failing, and if so how? Or do the data in fact deceive?"

Many development economists have debated the validity of the micro-macro paradox using quantitative analyses, but their studies have not systematically analysed how 'success' is defined and the data that make up the Bank's project-level ratings (see, for example, Subramanian and Ragan, 2008; Picciotto, 2009; Arndt, Jones, and Tarp, 2010; Howes et al., 2011). I wanted to look more closely at the question of "do the data in fact deceive?" from a different angle, by focusing on the details of the data upon which

this claim is based. I sought to investigate if this apparent contradiction is instead the result of the way that aid is assessed and how these assessments may be shaped by the context in which they are created and the professionals who determine them.

I was also motivated to pursue this topic due to my ongoing academic interest in how those with more radical viewpoints assess the system of international development. Based on theoretical and political arguments, rather than quantitative analyses, geographers and radical development economists, among others, have criticised the World Bank and similar organisations for their neoliberal viewpoints that describe poverty as a 'problem to be solved' rather than the result of neoliberal agendas and capitalist patterns of inequality (see, for example, Crush, 1995; Escobar, 1995; Rahmena and Bawtree, 1997; Peet and Hartwick, 2009; Mawdsley, 2007). Despite geographers' often valid criticisms of the World Bank, they have not produced a rigorous analysis of how the evaluation process at the Bank produces the concept of 'aid effectiveness' at different scales. 'Scale' is a significant geographical concept that, as I shall argue later, is a valuable way to analyse evaluation data and the production, measurement, and meaning of the concept of aid effectiveness.

1.3. Research objectives and questions

The objective in this dissertation is to work towards reshaping the aid effectiveness debate by exploring the context in which the Bank evaluates its work and how this institutional context shapes the Bank's data and evaluation reports. Given that the Bank's project-level ratings are critical in communicating the impact of aid to the clients of aid programmes and to the public, it is important to understand what these ratings measure and the institutional factors that shape them. The goal is not to criticise the Bank, but

rather to explore and articulate the data that the Bank's evaluations teams create. Academics and others analysing the impacts of aid often use these data uncritically rather than thinking carefully about what evaluations measure about the impacts of international development. I assert that understanding these institutional factors helps to explain the meaning of these data and their importance. The intersection between the professional practice of evaluation, the institutional spaces of evaluation, and the professionals who create the 'narrative' of aid effectiveness is the main focus of this analysis.

The second main objective is to provide a more comprehensive explanation of the concept of 'aid effectiveness' in practice. A significant issue in the analysis, as I noted above, is the 'micro-macro paradox,' which Mosley first articulated in 1986 and others have continued to analyse (Mosley, 1986; Subramanian and Ragan, 2008; Picciotto, 2009; Arndt, Jones, and Tarp, 2010; Howes et al., 2011; Tierney et al., 2011). The conventional economists, who write these studies, are generally trained in the field of neoclassical economics and often assume that there should be a connection between 'successful' projects and the economic growth of the recipient country. 'Successful' projects are often defined as having a high reported economic rate of return. I investigate what makes these projects 'successful' in the Bank's estimation and the context in which these projects are assessed and compared to the country-level results. The specific questions that are addressed and answered are the following:

- What is the context in which evaluation takes place at the World Bank and how does the role of evaluation within the Bank influence its evaluation reports?
- What are the evaluation processes at the Bank, how do they function at different scales, and what do evaluations measure?

- Who are the elite actors in the process of evaluation and how do they engage with and impact the results of evaluations in the Bank?

These questions have important implications for the understanding of the practice of evaluation and the data that are produced through these channels. In answering these questions, the regulatory role and power of evaluation and its elite actors, the theory and practice of evaluation, and the micro-macro paradox and the question of scale in the aid effectiveness literature are considered.

These research questions changed over time as the investigations challenged the initial assumptions. I began this project with several assumptions and hypotheses about aid effectiveness and evaluation: that evaluation is a straightforward quantitative exercise; that evaluators shape this process significantly; and that evaluation is a static discipline that generally has remained the same over time. The interviews with evaluators and the analyses of the evaluation documents challenged the original hypotheses and raised new issues, particularly about the role of evaluation within the Bank and how this role shapes the final evaluation products in the Bank, which became central to the analysis. Learning more about newer methodological approaches such as impact evaluation (Banerjee, 2007) and the institutional roles of evaluation within the Bank became important to understanding how 'aid effectiveness' is measured. This changed the questions to ones which more explicitly focused on the institutional factors that influence the professional practice of evaluation and its results.

These research objectives and questions are addressed in three steps. First, using interviews with Bank evaluation professionals, I provide concrete examples of how the World Bank measures its impact at different scales, and identify how certain institutional context and professionals shape the data. Secondly, using the full evaluation documents

and the data they use to justify the ratings, an analysis of what the Bank's evaluations actually measure in project-level and country-level evaluations is provided. Finally, drawing on both the interviews with evaluators and the case studies of evaluations at the project level and the country-level, I conduct a critical analysis of the approach and conclusions of the 'micro-macro paradox' debate in mainstream economics and then articulate a more comprehensive theory of how the concept of 'success' and 'aid effectiveness' are being produced through professional practices in the Bank.

As a result of focusing on the evaluation data, processes, and role of evaluation in the Bank, I do not make recommendations about how the Bank's projects or country-level work should be improved, but rather clarify how these processes and roles operate and how they influence the Bank's evaluations. Regardless of personal views on or experiences with the development industry, it is undeniable that organisations such as the World Bank create the dominant narrative (Escobar, 1995). Overall, I argue that the lack of attention to how this narrative is produced is an obstacle to better understanding the institutions which pursue international development and the work they complete in countries and communities around the world.

1.4. Summary of chapter contents

In this section, I summarise how the central questions about the production of aid effectiveness at different levels are addressed in this dissertation. There are eight chapters in total, ending with a conclusion that presents an overarching set of findings from the four empirical chapters and suggestions for further research. The literatures that are relevant to the questions and the methodology that is used to answer the questions are explained in chapters two and three, respectively. Combining the theory of economic

development with a more grounded understanding of how this definition is currently rendered through specific processes in the halls of an influential development organisation suggests that there is a more complex definition of 'aid effectiveness' in practice than whether or not it 'works.' The five diverse areas of literature from which the research questions emerged are discussed in chapter two: development economics, the critiques of international development and aid, institutional analysis, literature concerning evaluation as a professional practice, and the theory of monitoring and evaluation. The literature review helped me to contextualise and explain the different ways in which the problems and solutions of development have been approached, setting the framework for the empirical work.

The issues and the points of contention are detailed in the literature review (chapter 2), but here I briefly discuss the five areas of literature, the critical subgroups within these categories, and the importance of each area of literature to the overall dissertation. First, I explain the theories in development economics that form the goals of development and the changes in the way that economists have conceptualised the problems that development means to address over time (Rostow, 1960; Krugman, 1998; Mkandawire and Soludo, 1999; Birdsall, 2006; Sen, 1999; Lancaster, 2007; Stiglitz et al., 2009). Next, the critiques of aid are discussed, which range from criticisms of the results of aid to assessments of the potentially problematic core assumptions upon which it is based. This discussion begins with the literature that focuses its criticisms on the results of aid, including the micro-macro paradox literature (Mosley, 1986, 1987; Hansen and Tarp, 2000; Subramanian and Ragan, 2008; Doucouliagos and Paldam, 2009; Tierney et al., 2011; Howes et al., 2011) and then the bases for other criticisms of the results of aid are articulated (Alesina and Dollar, 2000). Thirdly, the literature that criticises the

assumptions and processes of aid is explored, especially the work by geographers and other social scientists who critically assess the narrative of development and poverty alleviation that the World Bank produces (Crush, 1995; Escobar, 1995; Rahmena and Bawtree, 1997; Mawdsley, 2007; McGregor, 2009; Peet and Hartwick, 2009). This discussion is followed by an explanation of the literature which focuses on the problematic way that, some claim, the Bank's research and evaluations conceptualise space and place, an issue which was, for some critics, clearly evident in the Bank's 2009 *World Development Report* (Harvey, 2009; Peck and Sheppard, 2010).

After considering the key concepts of development and the critiques of this practice, I turned to a set of literature that explains the role of evaluation and professional practices within institutions. Given that evaluation is a specific professional practice that exists in particular institutional contexts, it was critical to consider the literature about institutional theory and how institutions change in order to achieve particular ends (Kingston and Caballero, 2009; Power, 1999, 2003; Scott, 2001; Powell and DiMaggio, 1991). The role of evaluation, as a particular professional practice, and how it has evolved to serve institutionally-defined ends is also discussed (Christie, 2007; Gilbert and Vines, 2000; Grasso et al., 2003; Picciotto, 2009, 2011; World Bank, IEG, 2013, 2012). I then investigate the previous ethnographies of organisations that focus, in particular, on the role of elites (Latour, 1979; Burawoy et al., 2000; Strathern, 2000; Broad, 2006, 2007; Essex, 2008; Merry, 2011) in order to contextualise the role of evaluators, who are among the elite actors within development organisations. Finally, in order to explain the substance and the assumptions of evaluative practice as it exists in the Bank, I discuss the literature by both academic researchers and professional evaluators that assesses the theories of monitoring and evaluation (Bourguignon and

Sundberg, 2007; Barry, 2006; Cracknell, 2000, 1996; Rossi et al., 1999; Rebien, 1996). This area of literature includes an analysis of more recent methodologies, such as impact evaluation, that have influenced the goals of evaluation in the Bank (Banerjee and Duflo, 2009, 2011).

Chapter three explains the methodology and the data that were collected to address the research questions. The methodology is qualitative and is based primarily on expert interviews with World Bank staff (chapters four and five) and content analysis of original World Bank evaluation documents (chapters six and seven). In order to work towards reshaping the debate about 'aid effectiveness,' an interdisciplinary approach is followed which brings together several fields of study that are often not compared. In this dissertation, the research focuses on the *process* of evaluation at the World Bank, especially the logical framework and language in samples of evaluations, rather than the end result (i.e., ratings) of these evaluations. In the methodology chapter, I justify the choice of the World Bank as a case study and the process by which I selected the sample of evaluation documents used in the content analyses in chapters six and seven (Krippendorf, 2004; Flyvberg, 2006, 2011). The interview techniques and questions and methods for analysing the evaluation documents and the Bank's official evaluative frameworks at the project and country-level are also detailed in the methodology chapter.

Summary of Empirical Chapters, Chapters four to seven

To answer the key research questions, the empirical chapters complement and build on each other, beginning with a wide frame and scope of analysis (evaluation within the World Bank as an institution) and narrowing to content analyses of specific samples of evaluation reports undertaken within this institution. In chapters four and five,

I analyse the evaluation space at the Bank and how evaluators in the Bank create a particular narrative about the impacts of aid through their work. The final two chapters, chapters six and seven, present the results of the qualitative content analyses of the Bank's evaluation results at the project-level and country-level. By using both the interview responses from evaluators and content analysis of World Bank documents in these evaluations, I am able to compare and contrast the results of the process and the key actors' (evaluators') perceptions of the process (Clark, 1998).

In chapter four, the role of evaluation teams and how this role influences the narrative about aid effectiveness is discussed. I examine the institutional role of evaluation and how this influences the Independent Evaluation Group's (IEG) work, authority, and engagement with the places of development. Drawing on how evaluators explain their work, I assess how the way in which evaluators describe the role of IEG reflects the types of information that evaluators are 'allowed' to consider and how the institutional role of evaluation influences the resulting data that are created within the IEG. The way in which the role of evaluation within the Bank has changed over time as a response to external criticism of aid more generally is also discussed, as are the challenges to the current methodologies of evaluation as newer methodologies have developed in academia and professional practice.

In chapter five, the analysis focuses on the role of experts and the process of evaluation at the World Bank. I assess how evaluation professionals describe the processes, goals, and limitations of the World Bank's evaluation system and how evaluation professionals view themselves and the skills they use to evaluate the Bank's projects and programmes. I specifically consider the issue of scale and how evaluators describe the process of evaluation at different scales in the World Bank at the project

level and country level. I also gauge, based on the interviews, how evaluators view their work and what types of data and analysis are deemed as being most important from this 'inside' perspective. The evaluators' responses in these elite interviews helped to articulate the specific forces that shape this data that are related to the viewpoints and preferences of elite professionals and the required institutional processes rather than to a straight-forward empirical assessment of the Bank's work.

In chapter six, I analyse a sample of Implementation Completion Report Reviews (ICR Reviews) in the most common area of project assessment available (Agriculture and Agribusiness projects). Professional evaluators complete an ICR Review for each of the Bank's projects. ICR Reviews contain the authoritative ratings for the project and the rate of return data that Mosley and other economists have used for their aggregate analysis of aid effectiveness (Mosley, 1986). Using the full evaluation documents rather than only the ratings data, I examine the data and logic by which each of the ratings are construed at the project level in order to articulate how project-level evaluations at the Bank define and measure success. This allows for the articulation of the issues these ratings consider and how the Bank is currently defining aid effectiveness in practice at the project-level, what its logical framework is, and how these evaluations contribute to understanding the issue of attribution in practice.

Finally, in chapter seven, the question of scale and the relationship between the Bank's project-level ratings and its country-level evaluation reports is assessed. I use country-level evaluations from the region in which the Bank has completed the most country-level evaluations (Africa), which includes country-level evaluations from Kenya, the country that Mosley (1986) focused on in his original study on the 'micro-macro paradox.' I analyse the logical framework and the kinds of data used to determine the

Bank's impact on the country-level within the Bank's evaluations. This allows for the articulation of the implications of the Bank's evaluative framework, the methodologies for those researching the definition of aid effectiveness on different scales of aid, and the link between project-level and country-level outcomes. In the end, these findings, based on the actual language and logic of the Bank's evaluation documents rather than the distribution of the ratings contained in these documents, pose a challenge to the literature that asserts a 'micro-macro paradox' in aid results. The authors of these previous academic studies generally assess neither the Bank's many non-economic impacts nor the complexity of the relationship between the project ratings and the Bank's impact at the country level, viewing the ratings as though they are created in a vacuum (Mosley, 1986, 1987; Hansen and Tarp, 2000; Subramanian and Ragan, 2008; Doucouliagos and Paldam, 2009; Tierney et al., 2011; Howes et al., 2011).

Summary of Conclusions

In chapter eight, the findings reached in each of the chapters are brought together, providing a unified conclusion about aid effectiveness, the role of evaluation, and how evaluation results are 'produced' in the context of the World Bank. The conclusions contribute to the current theory of monitoring and evaluation and the aid effectiveness literature by explaining how particular aspects of evaluation's institutional role and processes shape the production of aid effectiveness in the Bank. This provides context for the previous literature's claims about the relationship between the results of aid at different scales and also demonstrates ways in which geographers could use their unique lenses to make important contributions to the process of interpreting the evaluation data created by development organisations. The findings have important theoretical

implications concerning the meaning of aid effectiveness and the professional practice of evaluation within the Bank and there are also critical implications for development policy and those who work to improve development outcomes using this data created within the Bank.

Considering the many implications of the role of evaluation and the processes of evaluation at different scales of aid, in the end, I highlight the need for more qualitative and comparative analysis to consider how the World Bank evaluates its projects and its country-level impact. In so doing, the research in this dissertation will make important contributions within human geography and other social science disciplines and will also provide guidance to practitioners and 'experts' designing evaluation systems and working within evaluation units in development organisations.

Chapter Two

Reshaping the aid effectiveness debate:

A review of the relevant literature

2.1. Introduction

The research questions of this dissertation are about the context in which evaluation takes place at the World Bank, the professional practice of evaluation, and the meaning of the concept of 'aid effectiveness' in practice. These questions emerged from important themes and approaches in many diverse areas of literature which are discussed in this chapter in five parts: development economics, the critiques of international development and the Bank's work, institutional analysis, ethnographies about evaluation as a professional practice, and the theory of monitoring and evaluation. The literature is itself highly interdisciplinary and, as a result, includes sources that view the concepts of 'development' and 'evaluation' in disparate and sometimes contradictory ways. Reviewing this literature, which is often informed by diverse theoretical positions, helped to define the research questions and the analytic framework of this dissertation, in which 'development' and 'evaluation' are analysed in many forms: as professional practices, as organisational forms, as data sets, as theories, and as narratives that aim to communicate certain concepts, including that of 'aid effectiveness.'

In this chapter, I critically examine the different literatures on these topics and demonstrate how, in sum, they helped to define the research questions and build the theoretical framework of the arguments in this thesis. Academic and policy researchers have published numerous studies on the topics of development organisations,

professional evaluation practices, and the concept of 'aid effectiveness,' but there are few studies that analyse how these areas of literature intersect. I counterpose the different arguments and explain how the research questions were shaped by the varying ways in which these literatures articulate the problems that development ideally should address, characterise the utility of development organisations, and reflect upon the measurement and evaluation of development work.

I begin by discussing the evolution of the theoretical framework of aid developed in the field of development economics and the critiques within development economics of these assumptions and practices. Next, the criticisms of international development and the Bank are reviewed, starting with the literature that questions the results of development and followed by an examination of the criticisms of the assumptions and processes of international development as a system. The focus then moves to the literature on development organisations such as the Bank and the ethnographies that examine the role of evaluation as a profession. In the summary at the end of this chapter, I clarify how the combination of the different approaches and issues that are raised in the literature shaped the research questions and the analytic processes in this dissertation.

2.2. Development economics and the practice of development

The research question about the role of development organisations and the influence of elite actors in evaluation emerged, in part, from a critical assessment of conventional development economics, which is rooted in the theories of neoclassical economics. This area of literature has shaped the practice of international aid and the goals against which aid projects and programmes are evaluated within the Bank (Woods, 2006). In this section, the key ideas of conventional development economics are

reviewed in a chronological way in order to explain two issues that shaped the research questions: an examination of the core theories of development economics and the extent to which the Bank's policies and evaluation models have challenged these assumptions over time.

It was important to explain how conventional development economists describe the goals of development practice because these goals have formed the standards against which the projects and programmes of development organisations such as the Bank are judged. The goals of development practice are rooted in modern development economics, which became a distinct practice after the Second World War when western nations formed the Bretton Woods institutions, including the Bank, to rebuild the economic world system after the war (Todaro and Smith, 2006). It is a relatively recent field, but one that is rooted in the longer neoliberal academic tradition of mercantilism and the notion of 'free markets' as articulated by political economists, including David Ricardo (1817) and Adam Smith (1776). Within modern development economics, 'modernisation theory' is a key issue and is based upon neoclassical economic assumptions. Modernisation theory posits that the main 'problem' to be solved is to be found in underdeveloped societies, which should be developed into 'modern' states through industrialisation and economic growth (Lancaster, 2007; Jomo and Reinert, 2005).

This literature motivated an important question to be addressed in this dissertation, specifically the extent to which the Bank's current evaluation frameworks reflect two central assumptions of conventional development economics: that industrialisation and growth are the goals of development and that the causes of economic growth are universally applicable rather than place-specific (Jomo and Reinert, 2005). In the early years of development as a distinct theory and practice, W. W. Rostow's (1960) work on

modernisation theory was particularly influential and remains important to understanding development practice. Based on an historical approach that embraced neoliberal economic principles as an alternative to Marx's account of the development of capitalist economies, Rostow argued that all countries pass through certain stages in order to become modern states and that international capital could encourage countries to advance to the 'take off stage' of rapid, self-sustaining economic growth (Rostow, 1960). As Rostow explained, "The take-off is the interval when the old blocks and resistances to steady growth are finally overcome. The forces for making economic progress, which yielded limited bursts and enclaves of modern activity, expand and come to dominate the society. Growth becomes its normal condition" (Rostow, 1960, 7).

A second research question that emerged from this area of literature is the degree to which the Bank's evaluations and its evaluators maintain that the impacts of aid can be proved through the use of quantitative measures and indicators. As modernisation theory was being developed, mathematical models such as the Harrod-Domar model, developed in 1946, served to 'prove' that various neoliberal approaches would lead to economic development, such as increasing the capital rate or decreasing the depreciation rate (Todaro and Smith, 2006). Other political scientists also provided influential analysis, arguing that there is a quantifiable link between economic modernisation and democratisation (Lipset, 1959; Huntington, 1968). Due to the importance of democratisation as a political goal for developing nations, these studies encouraged support for neoliberal approaches within governments and development organisations (Lipset, 1959; Huntington, 1968).

The theories expressed in these conventional development economics studies are important to the analysis of the language and logic of the Bank's evaluations in this

dissertation. The way in which the Bank's evaluators explain the goals and methods of development and the extent to which these responses reflect the key assumptions of conventional development economics are explored in chapter five. In chapters six and seven, which concern the language and logic of the Bank's written evaluations, I assess how the theories of conventional development economics are articulated in the Bank's current evaluations and the logic of the evaluations, including the extent to which the Bank's determinations rely on quantitative indicators and how this quantitative analysis is described.

2.2.1. Challenges within development economics

In this dissertation, I take a critical approach to development, exploring not 'if' it works but, rather, what the language and logic reveal about the goals of development and how development efforts are measured and evaluated within a powerful organisation. In order to develop this critical lens and analytical framework, more radical development economists' criticisms of the Bank's assumptions about the 'problems' that development should address and how the Bank should address them were important to consider. In particular, this area of literature influenced the research questions by bringing attention to two possible critical problems with the Bank's assumptions: how the Bank articulates the 'problems' it should solve through its work and the role of the state within the Bank's projects and programmes.

In part, these questions emerged from reading work by radical development economists who articulated 'dependency theory' such as Singer (1949) and Baran (1957) that provided an alternative to modernisation theory. Contrary to the assumptions of conventional development economics, the Prebisch–Singer hypothesis suggested that

affluent countries' consumption patterns were responsible for some of the economic problems in developing countries because the standard of life in affluent countries is based on developing countries' cheap resources and labour (Singer, 1949; Baran, 1957; Mkandawire and Soludo, 1999; Ferraro, 2008).

The radical authors in this area of literature bring attention to what they see as the problematic nature of the Bank's approach and criticise development organisations for the way they undermine governments through certain development approaches. These authors argue that the Bank's current development approach limits the role of the recipient state and does not recommend investing in social welfare programmes to support wide-spread human development. Based on these critiques and findings, reform and radical development economists articulated different approaches for the work of development organisations. Drawing on the work of John Maynard Keynes (1936), who argued that state intervention was necessary to mitigate the negative impacts of unpredictable economic cycles, some radical development economists in the 1960s put forth that developing nations should have a mixed economy with a role for government intervention during recessions, in opposition to the Bank's approach of limited government intervention (Fletcher, 1989). Radical economists' critiques had some policy implications: In the 1960s, some countries, particularly in South America, began to implement state-led economic development through import substitution industrialisation (ISI) measures. These new policies allowed these developing countries to direct the nature of their own development, but these countries generally ended this policy in the 1980s due to pressure from the World Bank and similar organisations (Ferraro, 2008).

As a result of reading the criticisms of the Bank developed in these radical economists' studies, the dissertation takes a critical approach to how the Bank's staff and

its official documents describe the Bank's role and the goals of its work. I wanted to know more about how the Bank's evaluations and its evaluators articulate the Bank's role and the role of the state in development programmes. The analysis in this dissertation specifically considers the way in which evaluators describe the role of the Bank and envision the role of the recipient state in the empirical chapters to follow, especially in the analysis of how the role of the recipient state is judged in the Bank's evaluations, a issue that is important to the analysis in chapters six and seven.

2.2.2. The Washington Consensus and the Bank today

The Bank's theoretical assumptions inform the standards it judges its work against. One of the key questions to be examined in this dissertation is the extent to which the Bank's work reflects changing ideas about what the goals of development practice should be. Many studies that have engaged in a critical analysis of the Bank's policies find that the Bank's approaches to development have continued to follow the 'free market' prescriptions described in section 2.2. above and, more recently, some variation of what has been called the 'Washington Consensus,' an approach that has its legacy in conventional development economics (Birdsall and Fukuyama, 2011; Lancaster, 2007; Williamson, 1990, 2000). In this dissertation, I assess the extent to which this still seems to be the case, based upon the Bank's own reports, and the ways in which the Bank may be refining its position.

Understanding the ideas of the Washington Consensus and the way in which these ideas influenced the Bank were particularly important to building the theoretical foundations of this dissertation. In 1990, John Williamson, a conventional development economist, articulated the tenets of the Washington Consensus, which consisted of ten

specific economic policy prescriptions such as macroeconomic stabilization, open economies for trade and investment, and the expansion of market forces (Williamson, 1990, 2000). The Washington Consensus formed the theoretical justification for the Bank's and other donors' adoption of structural adjustment programmes (SAP), a generic term to describe programmes which focused on limiting social spending and opening the market in recipient countries as a precondition for aid (Mkandawire and Soludo, 1999; Konadu-Agyemang, 2000). The SAP approach is based on the premise that recipient countries must restructure their economy to promote the ability of the free market to 'work' and that, in addition to financial reforms, countries must reduce various forms of government intervention (Mkandawire and Soludo, 1999). The conventional development economists who supported these policies predicted that inequality in developing countries would 'melt away' after economic development occurred. In accepting an SAP, recipient governments were required to reduce social spending and privatise state functions in order to reduce the government's intervention in the economy and allow free market policies to 'work' (Konadu-Agyemang, 2000). Despite the support for the SAPs and the Washington Consensus, these policies did not have the intended effect in many developing countries in the 1990s. As a result, economists from developing countries such as the authors featured in Mkandawire and Soludo's volume on African development (1999) emerged as vocal critics of the Bank's work.

The way that the Bank adjusted to these perceived policy failures is an important question that this dissertation addresses, because the Bank's evaluation practice grew, in part, as a way to justify the Bank's work in the face of these criticisms. A key research question became the extent to which evaluation reflects changes in the Bank's assumptions of and goals for its work. Birdsall and Fukuyama (2011) suggested that

these changes in the underlying assumptions of development and the Bank's criticism of the Washington Consensus mandates signals a 'post' Washington Consensus era, although there is some disagreement in the literature regarding how significant this change is (Rodrik, 2006).

The views of 'critical' development economists, who perhaps would not identify themselves as 'radical' but also are more critical of the tenets of neoclassical economists than the 'conventional' development economists described in this chapter, are instructive in understanding ways in which the Bank's standards may be changing. These critical development economists, such as Stiglitz et al. (2009) and Sen (1999), who are themselves conventional development economists, have articulated the problems with the previous ways in which conventional economists thought about and measured poverty. They have argued that, in some ways, there has been a shift towards recognising poverty alleviation and strengthening the role of states in the development process as important goals of aid (Sen, 1999; Collier, 2007). Sen (1999) proposed a new way to measure poverty with his Human Development Index (HDI), which includes indicators for aspects of poverty such as housing, food, and access to sanitation rather than just objective wages and other typical measures of poverty. The Stiglitz Report (2009) drew further attention to poverty alleviation as a key goal of development practice during the global economic crises that began in 2008-2009. Stiglitz and his colleagues (2009) advocated that development organisations should change their emphasis from measuring economic production to measuring the 'well-being' of people in recipient states. They make five general recommendations related to the goals of development, emphasising the multi-dimensional aspect of economic growth and poverty (Stiglitz et al., 2009). This work is important because it articulates a more comprehensive view of what poverty means and

the many ways in which it manifests itself, issues which may be important to the evaluation of aid.

The extent to which these 'new' aims and goals for aid have been incorporated into the Bank's evaluations is assessed throughout the empirical chapters of this thesis, especially in chapters six and seven, which assess the language with which the written evaluations explain the goals and shortcomings of the Bank's efforts. As a result of the arguments from more radical and critical development economists concerning the goals for development practice, this dissertation also assesses how the Bank's evaluators describe the goals for development and evaluation practice, particularly in chapter five which focuses on the 'art' of evaluation as described by the Bank's evaluators.

2.3. Critiques of the *results* of international development

The research questions that concern the evaluation data at the Bank and how they influence the meaning of 'success' in aid are drawn, in part, from conventional economists' criticisms of the Bank's reported results. The literature in this section focuses on the Bank's reported measurement of the results of aid such as the economic rate of return (ERR) of particular projects and national-level economic growth. Engaging with this literature informed the decision to focus on evaluation practices and influenced the research questions about the meaning of the concept of 'aid effectiveness' both in practice and in how the Bank currently determines this through its evaluative practices. A comparison of the Bank's rhetoric about the goals and the standards by which it judges its own work are important issues in this dissertation and were, in part, motivated by the studies and critiques discussed in this section. The studies described in this section also

influenced the decision to use the Bank's evaluation data in particular as a means to assess the meaning of the concept of 'aid effectiveness' in practice.

2.3.1. The 'micro-macro' paradox

One of the questions that is addressed most clearly in chapters six and seven is related to the issue of 'scale' in determining the success of aid and the relationship between project-level and country-level evaluations. This question materialised from the literature about the 'micro-macro paradox,' which instructed the decision to focus on evaluative practices and the nature of the data produced by the Bank in particular (Mosley, 1986, 1987; Hansen and Tarp, 2000; Subramanian and Ragan, 2008; Tierney et al., 2011). The strong criticisms of the results of aid from conventional development economists are important to this research questions because they suggest ways in which aid may fail even on its own terms, within the goals defined by conventional development economists. Furthermore, the fact that there are contradictory findings within the 'micro-macro paradox' literature motivated the focus in the analysis on the Bank's evaluation teams and how they create the data used in evaluations, rather than following the quantitative methodologies of conventional economists. As a conventional economist stated, case studies could "bring out the full richness of a national context as it influences aid effectiveness -- detail that gets lost in the numbers of quantitative analysis" (Roodman, 2007, 21).

There are several key studies in this area of literature that have defined the premises and the methods of the debate. In 1986, Mosley first described what he called the 'micro-macro paradox': that the impact of individual aid projects may be positive but that aid's aggregate impact is very small or even negative. In his original argument,

Mosley (1986) used World Bank data from 1961 to 1981 for projects in Kenya. In completing various regressions, Mosley found that there had not been significant economic growth, as other measures of growth such as savings rates and export growth had stayed the same, despite the fact that the World Bank rated the majority of its own projects as being 'successful.' By 'success' in this context, Mosley meant that the Bank reports a positive economic rate of return (ERR) for its projects and also references the high percentage of 'successful' project ratings that the Bank reports in its self-evaluations. In the 1990s, following Mosley's lead, Peter Boone (1994) published an influential article in *The Economist* which analyzed the growth patterns of ninety-six countries between 1971 and 1990 and concluded there was no relationship between increased aid and economic growth in recipient countries.

The decision to focus on the Bank's data collection and evaluative practices was influenced by the different conclusions that studies in this field have reached despite using the same evaluative data from the Bank. Mosley's study and the ones that followed it are often referred to as the 'aid effectiveness literature' (AEL). The main focus of the analysis in this literature has been on the extent of the impact of aid on economic growth. Subramanian and Ragan (2008), among other economists, also found that there is a micro-macro paradox and have used these findings as a justification for calling the aid system to be redesigned, usually drawing on neoliberal strategies and free-market reforms (Subramanian and Ragan, 2008). Hansen and Tarp (2000), on the other hand, concluded that aid does, indeed, promote growth. Arendt, Jones, and Tarp (2010) found that recipient countries' per capita growth rate increased by more than one percentage point per year in the long run and that aid stimulates investment (Arndt, Jones, and Tarp, 2010). Despite significant attention to the relationship between the Bank's project results

and country's economic growth, therefore, there remains uncertainty about aid and whether or not it achieves or can achieve the professed goals of donors. In completing a meta-analysis of the aid effectiveness literature, Doucouliagos and Paldam (2009) find that there is a tendency of economists to confirm the previous findings of the micro-macro paradox: "If an [Aid Effectiveness Literature] AEL researcher finds several results he/she will be reluctant to report those that suggest that aid causes harm. Rather, the most significantly positive result is likely to be selected as the key finding for a given aid effectiveness study" (Doucouliagos and Paldam, 2009, 456). The fact that scholars reached opposite conclusions while using the same types of evaluation data motivated the analytical framework of this dissertation, which aims to reconsider the aid effectiveness debate using a different approach.

The micro-macro paradox literature influenced the research questions and approach in all of the empirical chapters, particularly the focus in the questions in chapters seven on different 'scales' of aid as defined in the Bank's evaluations, which are different at the project- and country-level. The previous micro-macro paradox literature suggests that there is a positive relationship between the Bank's projects and economic growth in its clients' countries. There are many assumptions included in this claim, particularly that there should be a relationship between aid at different 'scales' and that the main goal of aid should be narrowly-defined economic growth. In this dissertation, I assess these assumptions in the micro-macro paradox research and the definition of 'success' and 'aid effectiveness' within the context of the Bank by taking a new approach: one which engages with the processes that form these data rather than only assessing the trends in the end results.

2.3.2. Other criticisms based on the results of aid

Conventional development economists have a variety of reasons for criticising aid. The critiques discussed in this section motivated the empirical analyses of the Bank's evaluation documents in chapters six and seven and the focus in the analysis in this dissertation on the difference between the goals of aid that the Bank articulates in these assessments and the issues by which evaluators judge the Bank's work in the evaluations.

First, in other results-based critiques, conventional development economists suggest ways that the Bank and other development organisations might be able to improve the way they disperse funds. David Dollar, a World Bank economist, is an influential contributor to this area of literature and, as Robin Broad (2007) argued, Dollar's conclusions about the way to optimise aid often influence the policies the Bank follows. Burnside and Dollar (2000) and other conventional economists allied with the Bank have put forth the 'conditional argument,' which posits that aid can lead to economic growth, but only when recipient governments have 'good' policies of free trade and fiscal responsibility (Broad, 2006; 2007). This argument has also influenced policy making in many donor organisations such as the World Bank, the Asian Development Bank, and the African Development Bank. All designed programmes with performance-based allocations which direct funding to countries with 'good' policies, by which they mean generally following neoliberal prescriptions that support a free market approach with limited state intervention (Broad, 2006; 2007).

Other critiques of the results of aid focus explicitly on the 'failure' of development organisations. Conventional economists (Kilby, 2000; Knack and Rahman, 2007) have found that donors fail to coordinate their efforts on a national scale, leading to smaller projects that typically are inefficient in the allocation of resources. This

phenomenon, referred to as 'donor fragmentation,' is based on the premise that the decline in project size and aid allocations undermines the effectiveness of aid and impedes the recipient government's ability to use aid well. In reviewing the data, these conventional development economists find that donors have given more to countries with the least need, despite attempts to change these patterns over the past several decades (Kilby, 2000; Knack and Rahman, 2007). Furthermore, conventional development economists have argued that aid may have a negative impact on governance by reducing savings that would otherwise build the financial capacity of the state (Radelet, 2006).

Next, the research question on the difference between the Bank's stated goals for its work and the actual language and logic by which it evaluates its projects emerged as a research question from the queries of conventional development economists (Milner and Tingley, 2010). As a result of the difference between what donors say and where they allocate funding, Alesina and Dollar argued that one reason development finance has been only somewhat successful at leading to growth and reducing poverty is that factors other than need determine aid allocation. They generally found that strategic and political considerations apparently drove donor decisions on recipients and the amounts given to them (Alesina and Dollar, 2000). The question of the aims and the intended goals of the Bank's funding at the project and country-level is important to the analysis of the evaluators' feedback, presented in chapters four and five.

Another question that conventional development economists have raised is the suggestion that other factors besides economic return may be more important indications of aid's effectiveness. Clemens et al. (2004) and other conventional economists argue that aid effectiveness may instead increase school enrolment or have other positive effects (Clemens et al., 2004). Other potentially important project-level factors that

conventional economists have identified as promoting positive returns include the participation of local stakeholders (Kingdom and Reddel, 2006). As Barder argues, "Because donors focus on just one dimension of poverty reduction (growth) they marginalize other legitimate objectives such as reducing chronic poverty or providing social services in countries that cannot otherwise afford them" (Barder, 2009, 1). This area of literature encouraged a consideration of the many types of non-economic indicators that are present in the Bank's evaluations, which are analysed in chapters six and seven, and an assessment of how evaluators explain that they can measure the Bank's success in reaching these goals, which is presented in chapter five.

Finally, the most drastic and fundamental criticisms of aid by conventional development economists were also important to consider. In the so-called 'dead aid debate,' economists (Moyo, 2009; Easterly, 2006) suggest that there are fundamental flaws in the international development system, calling for a complete rejection or reformulation of international aid as a system. In recent years, Dambisa Moyo (2009) and William Easterly (2006) have emerged as key proponents of the 'dead aid' conclusion. Their research speaks to a common question: why, despite such a massive investment in Africa and other economically disadvantaged regions of the world, are recipient countries are not seeing significant improvements in their growth rates? (Easterly, 2006). Moyo's (2009) suggested plan of action is rooted in free-market based solutions (bond trading, direct financial investment, and trade) and claims that the state should use the money earned through these investments, to finance its own education and healthcare programmes. In a similar vein, Collier argued that people within developing countries have experienced negligible income growth in the 1980s and 1990s (Collier, 2007). These stronger criticisms from within the field of conventional development economics

led me to focus the analysis in this dissertation on the way in which the Bank's staff define and justify the utility of its efforts in its publically-available evaluations, which is the focus of chapters six and seven.

2.4. Critiques of the *assumptions* and *processes* of international development

As opposed to the conventional economists discussed in the previous section that focus on the quantitative results of aid, the critiques of aid explored here focus on the assumptions and processes of international development (Crush, 1995; Escobar, 1995; Lawson, 2007). I wanted to explore the context in which evaluation happens and the 'unseen' factors that may influence the results. As a result of a review of the literature in this area, I focused on the institutional role of evaluation and the viewpoints of professional evaluators. The literature explored in this section, from diverse fields such as anthropology and development geography, helped to define the analytic framework of this dissertation by stressing the importance of understanding the Bank's 'narrative' and how it is 'produced' as a way to understand the Bank's evaluations. There are two distinct areas of criticism that shape how the issues of the 'production' of the Bank's narrative and 'scale' in aid effectiveness are analysed in this dissertation: the critique of the neoliberal assumptions and 'narrative' of development (Escobar, 1995; Crush, 1995; Rahnema and Bawtree, 1997; Lawson, 2007) and, secondly, the criticism from within economic geography of the Bank's New Economic Geography ('NEG') approach and the implications of how the Bank conceives of 'space' (Harvey, 2009).

There are two ways in which this literature, as an entity, shaped the research questions and approach in this dissertation. First, the approach of these studies informs the research, in which I focus on the internal logic of aid rather than the quantifiable

results it reports in its ratings data and how it is described, measured, and justified within the Bank. In particular, this literature motivated me to analyse the role of evaluation within the Bank as an institution and how this particular organisation's assumptions influences how they represent their work to the public. Secondly, the approach of these studies influenced the research questions in this dissertation that scrutinise the way in which geographical concepts such as space and place are described and included in the Bank's assessment of its own work. These areas of literature, much of it within the subfields of development geography and economic geography, are united by a common concern for what development means, the power it has to shape the narrative of development, how development organisations conceive of the spatial aspects of their work, and the assumptions upon which the practice of development is based (Lawson, 2007).

2.4.1. The narrative of development

The research questions about the way that the Bank's evaluation teams 'produce' particular concepts emerged from the approach in development geography and in post-structural critiques that assess, from a radical or at least 'critical' point of view, how development organisations promulgate certain ideas about development and aid (Crush, 1995; Escobar, 1995; Rahmena and Bawtree, 1997; McGregor, 2009; Peet and Hartwick, 2009). The authors of these 'post-structural' studies, generally anthropologists and human geographers by training, assess development as a dominant discourse supported by an established set of power relations that dictate the narrative and policy decisions of foreign aid (Escobar, 1995; Crush, 1995). These authors argue that organisations such as

the World Bank and International Monetary Fund (IMF) shape a certain view of what aid effectiveness means (McGregor, 2009).

A central question that this dissertation assesses is the meaning of 'success' in development as defined by the Bank, a question that emerged from post-structural analyses of aid. This required an analysis of the 'narrative' that is evident in the Bank's evaluations. To paraphrase Escobar (1995) and Crush (1995), who are key authorities on this subject, a narrative is a unified explanation of internal processes that is published by an organisation to communicate to the public and to its supporters the goals it has set for itself and the importance of its work in reaching those goals. Following the focus of the post-structural literature, the analysis in this dissertation focuses more on the language and the themes that development organisations, in this case the Bank, use to describe their work, and the power dynamics by which the elite actors, such as in this case evaluators in powerful institutions are the ones who are able 'to tell the truth' (Escobar, 1995). James Ferguson's (1994) research in Lesotho was one of the earlier studies on the topic. Ferguson identified the failure of programmes in terms of stated objectives and, so, Ferguson argued, development itself must be assessed, suggesting that the stated objectives were problematic, particularly in how they conceptualise the developing world. As a result of studies and approaches such as Ferguson's, the question of the type of knowledge that is valued in development is important to this research.

The way in which the Bank's evaluations articulate the 'problems' to be addressed by the Bank's projects and programmes was important to understand and articulate. Radical post-structural scholars interested in development (Lawson, 2007; Rahnema and Bawtree, 1997) are critical of 'problems' that organisations such as the Bank diagnose in the world and how these powerful organisations intervene in the places

of development. Similarly to the way that proponents of dependency theory (Ferraro, 2008) criticised the Bank, these scholars (Crush, 1995; Escobar, 1995; Rahnema and Bawtree, 1997) have argued that the way in which development organisations such as the Bank state their objective is problematic because it casts poor countries as the problem, rather than of 'overconsumption' (Illich, in Rahnema and Bawtree, 1997). In order to illustrate the problematic nature of the narrative of development, Wolfgang Sachs brought together articles from radical scholars who break down the assumptions of the key terms of development and provide alternative interpretations (Sachs ed, 1992). In defining 'poverty,' for example, Majid Rahnema concludes in this volume that "the time has come to look at poverty in a different way" and outlines an alternative vision of "convivial poverty" that is less concerned with "more having" than "more being" (Rahnema in Sachs ed, 1992, 172).

In the analysis of the Bank's narrative, the way in which the Bank describes itself within the narrative of development is a crucial issue. This analysis is influenced by the studies in this area of literature, which are critical of the fact that development organisations often articulate these problems as purely technical, whereas more radical development geographers argue that they are issues which should be resolved in the political arena (Rahnema and Bawtree, 1997). Importantly, post-structural studies often charge that the narratives of development are purposefully apolitical, a way to avoid the discussions about power differentials and politics about land, resources, and financial empowerment. Development geographers, who generally criticise aid from a radical perspective, have articulated that the 'problems' that development organisations seek to address are rooted in differences of class and power (Mosse, 2005; Harriss, 2002; Brigg, 2002; McGregor, 2009).

As a result of engaging with this literature, this dissertation uses the frame of analysing the development as a 'narrative' that these studies provide to critically assess the language and logic of the Bank's evaluation and how these documents and elite actors 'produce' a certain narrative about the impact of the Bank's work. In each empirical chapter, the analysis closely considers the words and logical structures by which the Bank's staff and publically-available documents justify and put forth certain explanations of the Bank's impact. The analysis of the narrative of development is closely related to the discussion of the power of evaluators in defining this narrative, an important question within this dissertation, particularly in chapter four.

The dissertation's research questions and 'case-study' approach also formed as a result of looking at studies by other geographers that are critical of development geography and post-structural analyses (Bebbington, 2003; Pieterse, 2000). As Vinay Gidwani (2002) argues in his study of irrigation projects in India, it is important for geographers to not to only criticize development, but also to use the unique lenses of the field to make a positive contribution to understanding the forces involved in development. Gidwani suggests that "our critiques must emerge not from stylised accounts of development but, instead, from concrete and case-by-case analyses of impacts" (Gidwani, 2002, 26).

The idea that the unique lenses and questions of geography and anthropology can help to reformulate the terms of the aid debate shaped the analysis in this dissertation by highlighting issues such as the narrative of evaluation, the language it uses, and how the evaluation narratives conceptualise of the spaces and places of development. By applying these lenses and frameworks in an analysis of the evaluation processes in the Bank, I use them in new ways which, I argue, can help to expand the debate about aid effectiveness.

The literature and studies discussed in this section were particularly important in defining the research because they shaped the decisions to directly interact with evaluation professionals in the Bank in order to analyse the narrative they construct, the focus of the analysis in chapters four and five.

2.4.2. Space, place, and scale

In assessing the Bank's narrative that is 'produced' in its evaluations, a critical question that arose from this literature is how the Bank's narrative includes the concepts of space, place, and scale. The work of economic geographers who have criticised the ways that the Bank conceives of scale in terms of development work inspired the questions about the issue of scale in evaluation. The analytical focus in this dissertation is explicitly on the evaluation processes as they exist in the Bank at two 'scales' of aid: the project and country-level, a focus which this area of literature motivated. These research questions and analytical approach emerged from the literature in development geography and economic geography that critically assesses the question of how elite actors, such as the evaluators who are the subjects of the study, include the spaces and places of development in their narrative about aid and how they engage with these spaces in their work.

To begin with, one of the key questions to emerge from the literature is the difference between how the Bank conceives of 'scale' in development and how more radical critics conceive of it. The approach to the analysis of the evaluation documents in chapters six and seven of this dissertation is based on critical economic geographers' view that, rather than being universal, knowledge is situated, place-specific, and grounded in the history and culture of those places (Lawson, 2007). Gillian Hart (2001)

provided the most succinct summary of the differences between how development economics and development geography analyse development in articulating a distinction "between 'big D' Development defined as a post-Second World War project of intervention in the 'third world' that emerged in the context of decolonization and the cold war, and 'little d' development or the development of capitalism as a geographically uneven, profoundly contradictory set of historical processes" (Hart, 2001, 650). The conventional economists described in section 2.2. generally follow the 'big D' approach in their analysis, whereas economic geographers tend to focus on the issues covered in the 'little d' approach.

The alternative conception of space and scale that more radical human geographers have articulated influenced the analysis in this dissertation by raising questions about the way in which elite actors, in this case the Bank's evaluators, engage with the places of development and how this influences the narrative they create. The work of development geographers that I draw upon here (Corbridge, 1986; Corbridge, Thrift, Martin, 1994; Cowen and Shenton, 1996; Peet and Hartwick, 2009) provides further insight into the problematic nature of the theories that conventional development economists and the Bank support by analysing the system of international aid through the lens of 'little d' development.

Secondly, as a result of reading this literature, I wanted to understand how evaluators' engagement with the places of development influences the narrative of evaluation. Massey (1994) and others have made important contributions in areas such as the connection between knowledge and power in development and the place-dependence of processes and the uneven workings of power. The global geographical span of evaluation is possible because of the advances in technology and travel that have

accompanied it. The ‘time-space compression’ noted by Harvey (1989) is reflected in the way evaluators describe the timing and global reach of evaluation (Massey, 1994, 149).

Finally, an important question that emerged from the literature in economic geography was how the Bank’s narrative reflects certain neoliberal principles about inequality and space in development practice. In particular, economic geographers, who write from both a conventional and more radical viewpoints, are critical of the Bank’s characterisation of ‘geography’ and ‘space’ in its explanation of development practice. Conventional economists’ and economic geographers’ competing conceptualisations of space and place in development work between was clearly defined in the publication of and reaction to the Bank’s 2009 *World Development Report*, entitled *Reshaping Economic Geography*.

The Bank’s 2009 World Development Report highlighted the Bank’s support of the concept of the ‘New Economic Geography’ (NEG), a theory proposed by conventional economist Paul Krugman (1998) to explain uneven geographic economic agglomeration. Krugman’s research was based on the premise that market mechanisms will increase geographical inequality rather than equally distribute economic activity, as opposed to what economic development models predict. Krugman argued that economic regions with the highest levels of production will be more profitable and will grow economically as more workers move from the ‘periphery’ to the ‘core.’ Rather than arguing that the best approach would be to distribute these regions evenly around the world, Krugman maintained that production and consumption should concentrate in a few locations (Krugman, 1998; Fujita and Krugman, 2004). Krugman’s argument had its precursor in the ideas about ‘spread’ and ‘backwash’ first articulated by Myrdal (1957), an institutional economist, in his discussion of ‘circular cumulative causation,’ and by

regional economists such as Perroux (1955) who developed the theory of 'growth poles,' arguing that economic growth often occurs around a 'pole' of a particular grouping of industry actors rather than equally across regional economies.

Drawing heavily on the theories of NEG, *Reshaping Economic Geography* (World Bank, 2009) served to highlight the divide been conventional development economists, who wrote the report, and radical economic geographers, who criticized the report's lack of attention to space and scale. The Report's authors argue that uneven geographic development can actually lead to economic growth and that economic inequalities will decrease as countries experience growth (Harvey, 2009; World Bank, 2009). NEG's proponents argue that there is a 'trade-off' between regional equity and national efficiency, which the World Bank endorsed to explain how uneven development increased during the decades of structural adjustment (Martin and Sunley, 2011; Wilson, 2011).

The research questions focus on the more radical economic geographers' criticisms of the Bank's Report and, thus, assesses more radical geographers' claims that the Bank espouses a development theory with 'universal' applications that does not engage with structures of power and class. A 2010 volume of the *Journal of Economic Geography* summed up many of these disagreements, with contributions from prominent economic geographers such as Uwe Deichmann, Jamie Peck, Eric Sheppard, and Gillian Hart (Deichmann et al., 2010; Lawson, 2010; Hart, 2010; Peck and Sheppard, 2010; Rodríguez-Pose, 2010). As Rodríguez-Pose explained, economic geographers' criticism falls into two categories: first, that the Report's main authors, who are conventional economists mainly employed by the Bank, generally overlook the work of geographers and did not include them in the panel of advisors and, secondly, that the Report was

simplistic in its view of fundamentally geographical issues such as space and place (Rodríguez-Pose, 2010). In an article on the same topic in a separate publication, David Harvey (2009) summed up how the Report depoliticises the issues related to inequality:

> The idea that market-led neoliberalism produces uneven geographical development is clear, but what the authors of this Report do not consider is how neoliberalism uses uneven geographical development as a means to promote the universality of its own world project, which has nothing to do with the well-being of the whole of humanity but everything to do with the enhancement of dominant forms of class power (Harvey, 2009, 1276).

This literature in economic geography as it relates to the Bank's acceptance of NEG raised questions about the Bank's articulation of the geographical concepts of scale, space, and place, which are at the centre of the analysis in each of the empirical chapters to follow. In analysing the evaluators' feedback in chapters four and five, I critically assess how they describe the places of development and the extent to which they explain development concepts as 'universal' realities rather than 'space specific.' Most specifically, the research questions for the written evaluations in chapters six and seven considers the way in which the language of Bank's publically-available materials examine the places of development and require an engagement with these places at different geographic scales.

2.5. Institutional Analysis: Evaluation and evaluators

The literature about institutions and professional roles within institutions informed the research questions about the context of the Bank's evaluations and the inner decision-making processes that influence the data that are produced. As a result of reading this literature, I wanted to be able to articulate the extent to which the institutionally-defined roles of evaluation and evaluators influence the evaluation results

and how they are produced. The literature is discussed here in two parts: the first part considers the role of evaluation and other auditing functions within institutions (Grasso et al., 2003; Kingston and Caballero, 2009; Power, 1999, 2003) and the second part reviews the literature about evaluators who are engaged in the profession of evaluation and the related organisational narratives (Gabriel, 2000; Strathern, 2000; Merry, 2011).

2.5.1. The role of the practice of evaluation within the Bank

An exploration of professional evaluators and the institutional influences on the information they produce became an important aim of this research. The literature surveyed in this section contextualised the interviews and the information that was collected about evaluation at the Bank within a wider body of research on development organisations and knowledge creation. More specifically, the literature in this area of scholarship directed how I analysed the role of evaluation and evaluation professionals in the production of aid effectiveness in chapters four and five.

The literature on institutional theory is drawn from a range of academic research and provided some context for the institutional forms of evaluation that the Bank currently has (Kingston and Caballero, 2009; Power, 1999, 2003). The specific details of the Bank's evaluation documents that are used in the sample for analysis in the empirical chapters are discussed in the next chapter on methodology; here the literature which assisted in the creation of a theoretical framework through which to analyse these Bank's documents are discussed.

The literature about institutional theory and the processes by which they change (Powell and DiMaggio, 1991; Power, 1999, 2003; Scott, 2001; Kingston and Caballero,

2009; Gertler, 2010) was useful in helping to identify the aspects of the role of the Bank's evaluation team (the Independent Evaluation Group, 'IEG') that may shape its evaluation results within a larger context of evaluation and other forms of auditing within institutions more generally. First, the connection between the institutional role of evaluation and its power is an important theoretical issue in this dissertation that the literature in this area helped to develop. Many scholars stress the normative mandate of evaluation: accountability, knowledge creation, and knowledge dissemination (Chelimsky and Shadish, 1998). Scott (2001) and other institutional theorists identify three bases of legitimacy within institutions and this literature helped me to understand IEG in terms of its relative organisational role. Barry's (2006) work on 'technological zones' was also important in this research in terms of understanding the types of roles that evaluation plays within the Bank within the wider context of institutions and businesses that have evaluation teams.

These sources on the practice of evaluation explain the link between evaluation as a professional practice and the institutional need for accountability more generally and the 'ritual' of evaluation practice within the Bank more specifically (Chelimsksy and Shadish, 1997; Power, 1999). Studies in this area of literature also highlight the importance of change within institutions, which became an important theme in chapter five which concerns the role of evaluation within the Bank over time. One dimension to consider is, as Kingston and Caballero's work (2009) explains, that institutional needs and knowledge can evolve within institutions as a result of 'deliberate institutional design' (Kingston and Caballero, 2009; Powell and DiMaggio, 1991).

In addition to the theoretical literature on institutional form and change, the research questions on the role of evaluation in particular emerged from the Bank's

internal explanatory materials about the history of evaluation. In order to analyse the 'narrative' the Bank produces, it was important to explore the historical reasons for the development of the role of evaluation within the Bank and the motivations for the particular structure of this department (Christie, 2007; Gilbert and Vines, 2000; Grasso et al., 2003; Picciotto, 2009, 2011; World Bank, 2005, 2013). Several important insights come from these internal narratives. The *World Bank Operations Evaluation Department: The First 30 Years* (2003), in particular, explains the chronology and the changing goals of development at the Bank and the institutional environment in which certain changes occurred (Grasso et al., 2003). IEG surveys among the World Bank staff are additional sources that were critical to understanding the purpose and the evolution of evaluation from the perspective of the Bank (IEG, 2013, 2012, 2011).

Comparing the way that these sources from the Bank described the role and function of evaluation with the way that the processes of change are described in the theoretical literature helped to articulate and to assess the Bank's role for evaluation within a wider scope of institutional functions (Kingston and Caballero, 2009; Powell and DiMaggio, 1991), an issue which is the focus of chapter four.

2.5.2. The role and influence of professional evaluators

The research questions on the role and the influence of evaluators, which are the specific focuses of chapter five, and, to some extent, chapter four, emerged from anthropologists' ethnographies of development professionals in other organisations (Strathern, 2000; Merry, 2011) and the literature about organisational narratives (Gabriel, 2000; Czarniawska, 1998). The literature on the ethnographies of development workers within organisations completed by anthropologists (Burawoy, 2000; Broad, 2006; Essex,

2007) and the role of auditing within organisations (Strathern, 2000; Burawoy et al., 2000; Strathern, 2000; Power, 2003) were particularly critical to the research questions.

The studies of the ethnography of elites in development organisations were critical to this analysis and put the role of evaluators as elites within development organisations within a wider professional context (Woods, 2006; Burawoy et al., 2000; Merry, 2011; Strathern, 2000; Broad, 2006, 2007; and Essex, 2012). The contributions in *Audit Cultures: Anthropological studies in accountability, ethics and the academy*, edited by Strathern (2000), particularly influenced how I developed the research questions about the nature of the work of evaluators and the extent to which the Bank's evaluators exhibit the traits of the professionals working in other institutionally-defined 'audit cultures' described in this volume. David Mosse's (2005) analysis of how development practices occur in large multilateral organisations and the relationship between policy and practice in development was also important. Two of Mosse's propositions particularly influenced the analysis of evaluation practices: "Development interventions are not driven by policy but by the exigencies of organisations and the need to maintain relationships" (Mosse, 2005, 16) and "'Success' and 'failure' are policy-oriented judgments that obscure project effects" (Mosse, 2005, 19). These propositions raised the question about what drives evaluators' determinations, which became important in the interview questions and analysis.

On a related note, Foucault's (1980) notions of power and knowledge informed the questions about how evaluators understand and communicate knowledge. Foucault's work helps to illustrate the distinction between what is considered 'legitimate knowledge,' such as the authoritative view of the Bank's evaluators in defining the official discourse, and 'local knowledge,' in this case that of the beneficiaries. As

Burawoy and others who have studied auditors within institutions have argued (Strathern, 2000; Merry, 2011), the roles of auditors often have the authority to determine the truth based on data. Although I found no studies that have investigated the work of evaluators, in particular, within the World Bank, some researchers have provided important analyses of structures and power within other development organisations such as the International Monetary Fund (Harper in Strathern, 2000; Woods, 2006). Broad's analysis of the Bank's research department, for example, shows how it reinforces a neoliberal development paradigm through the research it publishes (Broad, 2006). These studies proved to be important to developing the research questions about the nature of the role of evaluators within the Bank specifically and in the analysis of the language and logic that evaluators use in their work, which is central in all four of the empirical chapters.

Finally, in order to analyse the role of evaluators and the words they used to describe the results of development, I also had to think about the organisational demands and relationships within which evaluators' determinations are made. In particular, I had to understand and explain the process of creating organisational narratives in the Bank (Gabriel, 2000; Czarniawska, 1997, 1998). In analysing the expert interviews, I decided to consider not only the technical details of evaluators' work but also how evaluators saw themselves in the process. Organisational storytelling helped to explore this distinction as it offers some insight into this process, especially into the importance of analysing stories as a way to understand how wider organisational issues in the Bank are viewed (Gabriel, 2000). Within this area of ethnography, Latour's work on laboratory workers and their work in theory and practice and the distinction between them were particularly enlightening (Latour, 1979).

The literature about the influence of professional evaluators, auditors, and related professionals and the way in which their position and their mindset may influence the narrative of evaluation and the information collected about development projects and programmes influenced the research questions. Perhaps most importantly, I had to determine the extent to which the Bank's evaluators fill an 'authoritative' role with sufficient organisational power to determine and translate the results of aid. As a result, chapter five includes an explicit analysis of the way in which evaluators describe their abilities, professional role, and the importance of their work, in response to questions which were derived from the studies discussed in this section. I also consider the issues that were raised by the studies included in this section in the analysis of the Bank's written evaluations in chapters six and seven, particularly the questions that emerged from engaging with the literature on organisational narratives (Gabriel, 2000; Czarniawska, 1997, 1998).

2.6. The theory of monitoring and evaluation

The research questions about the way in which evaluators describe the process of evaluation were derived from the literature on monitoring and evaluation. This literature elucidated the technical aspects of evaluation in a comparative way and within the context of the wider professional field. This section, which reviews this area of literature, is divided into two parts. In the first part, the theoretical considerations in the literature are identified. These considerations raised questions about the logic and language of the Bank's evaluations and the interviews with evaluators. In the second part, the relevant literature in monitoring and evaluation is described. This literature clarified some of the

limitations of evaluation but identified some of the alternatives that might exist and that are explored in chapters six and seven.

2.6.1. Dissecting the logic and language of the evaluations

The literature on monitoring and evaluation informed the approach in the content analysis in all of the empirical chapters by explaining the logic and the assumptions of the processes used at organisations such as the Bank. The studies in this area of literature highlighted the considerations that are most important in determining which data evaluators use to explain their judgments on particular projects and programmes and why. In particular, the studies in this area suggested that the issues which are important to assess in order to understand an evaluation's logic are the logical framework, the use of indicators, and the issue of attribution.

First, subsequent to reading these studies, I wanted to be able to explain the Bank's logical framework in its evaluations. The log frame determines what kind of information evaluators use to base their decisions, what is reported in the Bank's evaluation reports, and what kind of data 'counts' (Maxwell, 1997; Cracknell, 1996, 2000; Grove and Zwi, 2008). Although there are many contextual factors to consider in selecting and designing an evaluation, the methodology and the type of framework chosen determines much of what the evaluation will reflect about the work to be evaluated (Conlin and Stirrat, 2008). In examining the Bank's logical framework, I also wanted to know the extent to which the Bank's project-level evaluations reflect international standards such as the standards provided by the OECD's Development Assistance Committee (DAC), the ECG (Evaluation Cooperation Group), and the United Nations Evaluation Group (UNEG) (IEG, Methodology; Cracknell, 2000).

Understanding these types of frameworks was important to consider in the overall picture of the Bank's evaluations.

The criticisms of certain frameworks which are discussed within the theory of monitoring and evaluation were also helpful in guiding my analysis. Many studies have argued that the log frame tends to hide the issues that require the greatest attention in development issues, such as indirect effects and unintended consequences (Grove and Zwi, 2008). As Grove and Zwi (2008, 71) articulate, a project may be judged to complete its outputs within this framework, but have unintended consequences or outcomes due to the fact that quantitatively driven evaluations minimise issues related to non-economic activities such as governance. These observations influenced the research questions posed to evaluators in the Bank, which focused on eliciting responses about the 'unseen' issues that may not be captured by the Bank's current framework.

In the process of analysing the Bank's log frame, the literature on the theory of monitoring and evaluation also motivated me to explore the processes of the Bank's evaluation and to compare them to other examples within the wider profession (Bourguignon and Sundberg, 2007; Barry, 2006; Cracknell, 2000; Rossi et al., 1999). Materials from professional organisations such as the American Evaluation Society (AES) explained the rise of evaluation in development as a professional field and the formation of organizations such as the Evaluation Network and the Evaluation Research Society. Rebien's (1996) review of evaluation theory and practice clarified the wider field of which the Bank's evaluation team is a part. It was also important to articulate the use of evaluation within development organisations such as the Bank as opposed to evaluation functions in other businesses and organisations. This literature defined distinctions between business evaluation methodologies, such as the New Public

Management approach that is associated with performance measures, and those employed in community organisations (Blalock, 1999). These studies provided context for the analysis of the Bank's logical framework in chapters six and seven.

Secondly, an important objective of this dissertation is to break down the sometimes 'jargon-filled' language of evaluations in order to articulate the implications of certain frameworks and words in this context. This area of literature explained what the use of particular evaluation approaches and indicators reflects about the nature of the determinations created within institutions such as the Bank. In this way, this literature helps to uncover what aspects of development projects the Bank's evaluation are designed to measure (Merry, 2011). In considering the Bank's work, the literature about the efforts of development organisations to reform their practices through public declarations also raised questions about the utility of indicators in the wider aid community and the need for targets to be reached in areas such as poverty alleviation (Conlin and Stirrat, 2008). By analysing the Bank's evaluations and the language used in them, the aim was to articulate how, exactly, such indicators are used in context. This literature raised an important question for the expert interviews and the analysis of the extent to which evaluators support the Poverty Reduction Strategy Paper (PRSP) approach which has required increased data collection and encouraged client countries to develop monitoring and evaluation systems to hold governments accountable by providing reliable data (Holvoet and Renard, 2007).

Finally, one of the main research questions in this dissertation concerns the relationship between the Bank's work and its project-level and country-level results, concepts referred to in the literature as 'causality' and 'attribution.' I wanted to understand how the theory of monitoring and evaluation explained this connection,

which is a key assumption of the micro-macro paradox research (Mosley, 1986). Previous studies on the theory of attribution in the context of international aid generally focus on the 'black box' between a project design and its impacts (Iverson, 2003; Bourguignon and Sundberg, 2007) and the institutional, political, and other reasons for average project success in some countries rather than others (Dollar and Levin, 2006). Although the micro-macro paradox assumes that successful projects have a clearly defined relationship, questions remain in the theoretical literature about the way in which the impacts of aid can be attributed to a particular project, the timing of evaluation, and the qualitative and quantitative methods that can be used (Iverson, 2003; Bourguignon and Sundberg, 2007; Cracknell, 2000; Rebien, 1996). I wanted to explore these questions. As Cracknell stated (2000, 114), "it has to be admitted that there is a degree of unreality about the crude assumption 'if a then b' that constitutes the underlying logic of the system." This literature on monitoring and evaluation informed how I analysed the complicated nature of these relationships in a more critical way that was grounded in evaluation as it is actually conducted. This literature was particularly helpful in designing and implementing the analysis of the Bank's written evaluations, which is carried out in chapters six and seven.

2.6.2. Limitations of the Bank's evaluation system

The studies on the theory of monitoring and evaluation illuminated the possible benefits of alternative evaluation approaches. The relevance of alternative frameworks and evaluation methodologies is assessed in the discussion of the chosen logical framework for the Bank's evaluations in chapters six and seven. One possible alternative, for example, is the 'theory based evaluation' formats that use more qualitative data

(Weiss, 1997; Pawson, 2003). Another model is the 'most significant change' theory which traces, through interviews, the most critical changes from the perspective of project recipients (Davies and Dart, 2005).

In the course of completing the interviews, the notion of 'best practice' in current evaluation theory became an important issue to consider while analysing the Bank's current evaluation practices. Methodologically, an important theoretical change in the field of conventional development economics and evaluation has been the increased use of a methodology called 'impact evaluation.' Impact evaluation uses randomised control trials to determine what would have happened in the absence of the project or programme. Modelled after medical trials for pharmaceutical products, impact evaluation creates a laboratory environment as much as is possible in the real world. In completing impact evaluations, evaluators set up experiments which intend to compare the quantitative outcomes of the beneficiaries of the project and a control group in which a particular project or program is not implemented (Banerjee and Duflo, 2009, 2011; Banerjee, 200). Studies based on the comparison of project outcomes between two groups have been undertaken, showing, through data collection and analysis, what would have happened to these beneficiaries if the Bank had not chosen to implement a particular project or programme there (Banerjee and Duflo, 2009, 2011; World Bank DIME, n.d.). I wanted to assess how the increased use of impact evaluation at the Bank (World Bank DIME, n.d.) may influence the idea of what the goal of evaluation is and this is among the issues addressed in chapter four, which focuses on the role of evaluation in the Bank.

A final issue that this literature raised that helped to break down the logic of the Bank's evaluations is the question of the extent to which the voices of beneficiaries are included in the evaluations of development organisations, an issue that is assessed most

clearly in chapter six. Some argue that the Bank's focus on performance measurement and monitoring serves to undermine the empowerment of local communities and is politically motivated. The problem is that by excluding alternative voices, technocratic approaches to evaluation work to depoliticize public policy decision making (Schwandt, 2009). Evaluation is often seen as a 'neutral' instrument because it delivers evidence-based information but, as a result of not considering the sources of information, these instruments do not capture local views. These participatory approaches to evaluation advocate for including local ownership, empowerment, and programme improvement within the evaluation mechanisms (Cousins and Chouinard, 2012).

Overall, this literature on the limitations and alternate approaches in evaluation shaped the research questions and the analysis of the Bank's written evaluation documents in chapters six and seven by demonstrating the choices that the Bank has made in designing its current evaluation system. These choices were important to consider in assessing the final narrative and also shed light upon the issues that the Bank's current system may not measure or consider to be important within its current evaluation instrument.

2.7. Summary

In sum, the many ways in which the issues of 'development' and 'evaluation' are analysed in these diverse literatures informed the research questions and the analytical framework in this dissertation. As a result of engaging with these sources, I analyse these concepts from different points of view: as theoretical positions, professional practices, data sets, and as organisational narratives. This varied analysis led to an exploration of the specific technical processes, institutional goals, and the manifestation of certain

assumptions within the Bank. In particular, the literature raised important questions that are important in completing the analysis related to: the inherent problems in the development model and the assumptions upon which it is based; the differences between the Bank's official narrative and how it evaluates its own programs; the role of evaluation within the institutional context of the Bank and how this context may influence the end results of evaluations; and the issues of power, knowledge, and political engagement that may be omitted from the narrative of aid in the Bank's evaluations and what makes its projects and programmes 'successful' in its own estimation.

First, in order to complete the analysis on the role of evaluation and the assumptions of evaluators in chapters four and five, it was important to articulate and understand the assumptions upon which the system of international development is based in a critical way. The literature provides both substance and context. In particular, the differences between how conventional development economists (Rostow, 1960; Alesina and Dollar, 2000) and radical scholars (Lawson, 2007) conceptualise the problems which development should address inspired the research questions about the Bank's theoretical assumptions and its resulting evaluation processes.

Next, this diverse literature shaped the research questions on the practices of evaluation and the role of evaluators in chapters four and five. The 'narrative' that evaluation creates about issues such as 'success' inspired the critical research questions about the language and logic of evaluation within a particular institutional narrative. The combination of institutional theory (Scott, 2001; Barry, 2001) and the literature on organisational narratives (Gabriel, 2000; Czarniawska, 1998) helped to explain the ways in which certain concepts are defined through institutional processes and why evaluators' viewpoints are important in the overall understanding of the Bank's evaluation data.

Finally, the literature contextualised the complex and often jargon-filled evaluations published by the Bank, explaining the implications of using certain frameworks and how they shape the data and the narrative in these assessments. The literature on monitoring and evaluation provided insight into these critiques and the official language of the Bank in order to critically assess the logical framework, the use of indicators, and the issue of attribution as they are developed in these documents (Holvoet and Renard, 2007; Merry, 2011; Banerjee and Duflo, 2009, 2011). This literature, in conjunction with the literature on the roles and goals of evaluation from the literature cited above, defined and gave substance to the content analysis of the Bank's written evaluations in chapters six and seven. In the next chapter, I show which methodological strategies I adopted to address these questions.

Chapter Three

Analysing the 'window' of the World Bank's evaluation data: Methodology and theoretical considerations

3.1. Introduction

In this chapter, the methodology for answering the research questions and the decisions involved in creating and executing the research plan is detailed. After laying out the reasons for the adoption of a qualitative methodology and a case study approach, I clarify the strategy for the expert interviews and review the relevant theoretical issues (Rose, 1997). Finally, the selection process for the samples of the evaluation documents used for the content analysis, the process for analysing these reports, and the theoretical considerations important in this process are explained.

3.2. Qualitative methodology

The literature on the merits of particular approaches in different geographical subfields suggested that qualitative methodologies, such as the expert interviews and discourse analysis in this dissertation, would be better suited to the research questions which seek to explicate institutional decision-making processes and the assumptions behind them (Lawson, 2007; Hart, 2001). From the many types of qualitative methodologies available, expert interviews and discourse analysis in particular were employed because these methodologies are best suited to the questions of this dissertation, which concern the nature of the Bank's evaluation process and the logic and language of its evaluation products. I chose these approaches because I wanted to know, in the first instance, the process of evaluation from the perspective of those doing it in

order to understand what the full-picture may be and what the 'behind the scenes' considerations are within the evaluation process. I also wanted to understand how evaluators describe and justify the results of the Bank's work in their written reports, to put the ratings data used in previous studies within the context in which they are created. It is important to establish that, although the research is primarily reliant on elite interviews and the analysis of written evaluations in different chapters, these are complementary methodologies by which to analyse and view these decision-making processes. In the empirical chapters and in the conclusion of this dissertation, the analysis of the written evaluations that the Bank's evaluators produce is compared to how they described these processes in the expert interviews. Being able to compare the results of the discourse analyses of the Bank's written evaluation to the evaluators' feedback in the interviews helped to confirm the findings presented in this dissertation (Clark, 1998).

The use of qualitative methods, in addition to providing an appropriate approach to the research questions, also responds to calls within economics, geography, and other social science disciplines for the use of innovative approaches to analyse the impacts of international development. Conventional development economists, who generally use almost exclusively quantitative methodologies, have criticised the predominantly quantitative aid effectiveness literature because these studies have produced conflicting and confusing results (Clemens et al., 2004; Doucouliagos and Paldam, 2009). As Clemens, et al. quipped, "Wearied readers of this literature would be right to wonder what produces diverse findings from apparently the same aid and growth data" (Clemens et al., 2004, 1). Some economists have suggested that using content analysis, such as the approach used in this study, would be useful to clarifying studies that rely solely on aggregated country-level data (Roodman, 2007).

3.3. Case study approach

In this dissertation, the focus of the analysis is upon a case study of the World Bank and its evaluation systems, processes, and staff. In this section, I explain why using a case study approach was appropriate and then clarify why the World Bank was selected, in particular, as this case study. Importantly, the case study provides a unified context for the analysis of the evaluation results. The case study approach has been defined as an "in-depth exploration from multiple perspectives of the complexity and uniqueness of a particular project, policy, institution, program or system in a 'real life' context" (Thomas, quoting Simons, 2011, 512). Using the Bank as a case study provides an analytical frame to ground the research in a particular setting and allows the comparison of these diverse evaluation documents and expert interviews within a common context (Thomas, 2011).

Within the field of international development, there are many different case studies that could have been selected, based on different factors, such as: evaluations of different types of aid, different locations of aid, or different institutional settings. I decided it was important to focus on a particular institutional context due to the fact that evaluation practices and methodologies vary between different institutional settings. In selecting an institutional context, there are two reasons that the World Bank was an appropriate case study. First, as I found in my attendance at the Bank's evaluation conference, the Bank has the largest evaluation team and most intensive evaluation process in international development organisations (Yannias, 2013a). The Bank is also a suitable case study because the original research on the micro-macro paradox focused on World Bank data (Mosley, 1986).

The decision to use a case study approach also responds to calls within development studies to consider the issue of scale and how case studies can be brought into a theoretical framework and scaled up. Geographers working on topics related to development have either worked on specific case studies at the exclusion of engaging in the wider debate or have joined the mainstream debate but have often lost their unique lens (Gidwani, 2002). Although detailed analyses of particular development approaches in specific geographic spaces is critical, many scholars have argued that it is also necessary for geographers and other social scientists to build a theory about the implications of development practices more generally (Bebbington, 2003).

3.4. Elite interviews

Elite interviews were critical to uncovering the relevant information to answer the questions about decision-making processes and to identify the most important factors that evaluators within the context of the case study use to create their data. In order to better understand this process from an inside perspective, I interviewed evaluation professionals who either currently work or formerly worked as evaluation professionals in the World Bank and present the results of these interviews in chapters four and five. The interviews with academic experts in the topic of development are important resources in the analysis of the role of evaluators and the process of evaluation. In this section, I describe: how the interviewees were selected and how the interviews were conducted, the interview questions and analytical framework, and the theoretical considerations for these expert interviews.

3.4.1. Selecting and conducting the interviews

The interviews were conducted with evaluation professionals rather than other actors involved in international development, such as the recipients of aid, because the dissertation questions specifically concern the professional practice of evaluation. In terms of the selection criteria, the academic researchers I interviewed had to be experts on the issue of development and have completed research that required them to work within development institutions or interview development professionals. The World Bank professionals I interviewed had to currently work or formerly work as evaluators in the World Bank. Most of the professionals included as interviewees currently work or formerly worked in the Independent Evaluation Group (IEG), the World Bank's formal evaluation group, but a several evaluators currently employed in other evaluation units of the Bank were also interviewed in order to have a wider sense of Bank's overall evaluation space. I attempted to speak with as many World Bank evaluators as was possible, given the relatively small size of the unit and the difficulties that were experienced in accessing these professionals.

Twenty-five interviews with academic and evaluation professionals were conducted in three rounds due to the availability of the academic experts and evaluators who responded to the interview requests (please see Appendix 1 for a summary of the interviews conducted). Although I had planned to complete all of the interviews at once, it was, in retrospect, useful to have time between the rounds of interviews because it allowed for a consideration of certain key issues and the refinement of the language in the questions. In February 2013, before interviewing the World Bank evaluators, five interviews with academic experts were conducted. These academic experts had completed research about the inner workings of development organisations such as the

Bank and some had experience working as consultants in the Bank or similar institutions. Given the international and far-reaching geographic location of these academic experts, these interviews were conducted via Skype, an online conferencing system that includes video capabilities. All of the academic experts included as interviewees are native English speakers and are employed in English-speaking academic institutions. I employed a conversationalist approach these interviews and asked open-ended questions not only about their experience conducting interviews and research with development professionals, but also about the state of evaluation research and which questions would be helpful to clarify in this dissertation. These interviews helped to direct the ongoing literature review, focus the research questions, and refine the approach in the subsequent interviews with World Bank evaluators. These interviews also clarified the research questions that could make an important contribution to the current research on aid effectiveness and evaluation in international development in practice.

In the months following the interviews with the academic experts, I began the interviews with World Bank evaluators. I also completed the interviews with evaluators exclusively in English and all of the evaluators included as interviewees are either native English or fluent English speakers. Given the status of these elite professionals and their demanding schedules, gaining access to evaluators proved to be somewhat difficult and it took several months to arrange the first round of interviews. To begin the process, I compiled a list of evaluators through the Bank's website and the names of evaluators that appear in the Bank's evaluation reports. I began contacting individual interviewees by email in late January 2013 and, in some cases, spoke with them by phone if they requested further information about confidentiality or the purpose of the proposed interview. In the introductory email message, I fully disclosed my status as a DPhil

researcher and honestly described the purpose of the research (McDowell, 1998). In the email, the research was described as follows: "My research focuses on the role of evaluation units in the development industry and the processes by which aid effectiveness is measured. I am interested in interviewing you in order to learn: your views on the function of evaluation and research in the development industry, your perspective on how evaluation may be changing, and how you and your team members work with professionals in different roles at the World Bank and at other development agencies" (Yannias, 2013b).

Using online searches and a review of the stated authorship of evaluation documents published by the Bank, I located current contact information for forty-five evaluators in total, but only twenty-five of those who were contacted responded to the email. Five evaluators who returned the email did not consent to be interviewed, providing different explanations: that their travel schedule did not allow, they were moving into a different department and being relocated abroad, or they work in administrative tasks rather than on-the-ground evaluation and did not feel they could comment on the issues relevant to the dissertation. The reason why many evaluators failed to respond to the introductory email cannot be definitively stated. A few evaluators raised the question of access in the interviews, noting that there is hesitancy among Bank evaluators to reply to interview requests from academics due to critical studies in the past. Of course, the interviewees cannot be expected to accurately explain their colleagues' decisions to not respond to the interview request, but this observation suggests a rationale for those Bank evaluators who chose not to participate in the interviews.

Evaluators were often concerned about confidentiality before agreeing to an interview for two reasons. First, there was a high risk that they may be identified in their responses as there are only approximately one hundred and twenty professional evaluators in the Bank. Secondly, there is significant public scrutiny about the issue of evaluation both between different teams within the Bank and in the wider community of development professionals. Given the valid concerns and the reluctance of the evaluators to be interviewed otherwise, I guaranteed anonymity in the interviews as many previous researchers have guaranteed in their research on the Bank and other development organisations (Woods, 2006; Broad, 2007).

The second round of interviews occurred from March to May 2013 with the first group of evaluator interviewees. These twelve interviews were conducted in Washington, D.C., at or near the Bank's main headquarters in the downtown area at 1818 H Street N.W. I completed two of these interviews by video conference (Skype) due to the interviewees being in transit and unable to commit to a specific interview time in Washington, D.C. I then began to analyse the second round of interviews and clarified a few of the questions for the later round of interviews, although the topic of the questions remained the same in both the spring and early autumn of 2013. I completed the third round of interviews in September 2013 with a different group of eight evaluators; these interviews were also conducted in Washington, D.C. at or near the Bank's main headquarters in the downtown area.

In addition to these two rounds of interviews with evaluators who work at the Bank, I also attended the World Bank's Evaluation Week in March 2013, which is a conference in which professional evaluators, from both the Bank and other institutions, present on issues that are important to the evaluation field. I was alerted to this

conference and granted access by one of the interviewees who suggested that I attend the conference before our interview. I took detailed notes throughout the conference and also engaged in informal discussions with the evaluators during the sessions. I included these observations from the informal discussions in the detailed conference notes that became a useful source in my later analysis. Attending this conference was an asset to this research and provided insight regarding the concerns and issues that evaluators consider within their professional circles and how they view the goal and the challenges of their work (Yannias, 2013a).

3.4.2. Considering the interview environment

I closely considered the environment of the interview and the status of my own knowledge about the 'insider' evaluation process at the Bank in making the arrangements for these interviews. I prepared for the interviews with the Bank's evaluators by researching the interviewees' background, job description, and work experience using their online publications and evaluation reports on which they are a named contributor. The online biographies of the interviewees on the Bank's website and websites such as LinkedIn provided introductory background information about the interviewees' educational background and work experience. As I organised the material about the various groups and work these professionals had completed, I learnt terms and acronyms common to the language of evaluation in the Bank, which allowed for more natural discussions during the interviews. In the interview itself, I presented the research not as a critique of the evaluation system but rather an inquiry into the way the evaluators saw their own career and role within the Bank, to "understand the world from your point of view" (Sabot, 1999, 329). From the beginning, I clarified that I was not working 'on'

development professionals, or deceiving them in order to get a critical angle on their work, but rather working 'with' them in an interpretative community to better understand their profession and their work.

An important consideration was the location of the interview. At the advice of the academic interviewees, I asked the interviewees to decide where the interview would be held in order to make them as comfortable as possible as they explained their work to an outsider. Most evaluators preferred to meet in one of the several cafes located within the Bank rather than in their office and three interviewees requested that we meet outside the Bank at a café near the main headquarters. I observed that evaluators who were interviewed in more public places rather than their offices were less distracted and more likely to provide me with more time in the interview. Of those that opted to hold the interview in their office, however, most of them were careful to close the door and block incoming phone calls, so we did have substantial privacy in the interview. Overall, in my estimation, the personality of the evaluator, rather than the location of the interview, determined how open and detailed the interviewees were in their replies to the questions.

A final consideration about the 'environment' of the interview was the decision to make an audio recording, with the consent of the interviewees. During the interviews with both the academics and the World Bank evaluators, I used a tape recorder so that the interview conversation would flow without me taking notes and, more importantly, so that I could capture the specific words and phrases that evaluators use to describe their work and role within the Bank. Twenty-four out of twenty-five interviewees agreed to allow me to audio record the interview. One interviewee refused to allow me to record the interview; they were not specific in their explanation for their refusal, only noting their concerns about the confidentiality of an audio file.

3.4.3. Interview questions

In this section, I describe the questions I asked the academic experts and the World Bank professionals and the reasons I had for asking them these questions. In the interviews with evaluators, I engaged in 'close dialogue' (Clark, 1998) and, to the greatest extent possible, positioned myself as a shared interpreter of their work. This meant that, rather than asking evaluators primarily about the specifics of the methodologies by which the Bank evaluates particular projects, I asked them for their views on data and measurement and what they thought of as their role within the Bank. I made several crucial decisions about the interview questions based on the feedback from academic experts in the interviews I completed before leaving for Washington, D.C.

The interviews were semi-structured, but, at the advice of several academic experts, I asked the questions in an 'open-ended' way that encouraged the evaluators to describe their opinions of their roles and activities. For example, I asked "what do you see and the goal and importance of evaluation?" rather than for an explanation of the common steps they take in their work. This was because I was interested not only in what the evaluators said but, importantly, the words and logic by which they described these issues (Mikecz, 2012). While I had a list of questions that I asked the interviewees, I generally took a conversationalist approach and, although I followed the same set of questions in the interviews, the interview questions were presented as themes rather than as set questions, encouraging evaluators to tell their own stories about their 'lived world' (Kvale, 1983, 1996). Overall, the goal was to understand the process from evaluators' perspective, to get their 'field' stories and be able to better envision the process from beginning to end, to put the written reports in context (Kvale, 1983, 1996).

The questions I asked the academics were different from the evaluators, due to the different aims of these two types of interviews. The academics were experts in the field of development and work on issues related to the practice of development, in particular the roles of those within development organisations. Although the academics were less familiar with the role of evaluators within the Bank specifically, they offered insight into how to speak to and what to expect in interviews with professionals working in environments such as the Bank. I asked the academics to provide candid feedback on the interview questions and overall research topic. Their feedback provided important insight into oversights with the original group of questions and suggested important theoretical contributions that this research could potentially make. I also asked the academic experts to describe their experiences of interviewing professionals in a setting such as the World Bank and what would be important for me to know in engaging in a research project that involved interviewing professionals at a powerful organisation such as the Bank.

A full list of the interview questions I asked the evaluators can be found in Appendix 2 at the end of this dissertation. In the interviews with the evaluators at the Bank, I asked a series of questions divided into three themes, covering the issues of both the organisational role of evaluation and the role of evaluators in the process of evaluation. The evaluators' responses to these questions inform the discussion in chapters four and five. The first group of questions focused on what evaluators do within their role and their perspective on that work. The evaluators were asked to explain their opinion of the evaluation process, to reflect on what they viewed as the most important skills that someone would need to do their job, and the reasons they were motivated to pursue their career in evaluation. This first group of questions served as an 'icebreaker,' because the evaluators felt comfortable talking about their careers and also helped me to understand

what evaluators view as the goal and limitations of evaluation as a practice. The second group of questions focused on how evaluators work with other groups within the Bank and how they viewed the role of evaluation in informing the Bank's work. This area of questions helped me to understand evaluators' networks, how they think about evaluation as a unit within the context of the World Bank, and how they describe their role and its authority. The final group of questions focused on how evaluators use data in their work, to what extent they believed that effectiveness could be measured, and if evaluation methodologies may be changing in the Bank. These questions helped me to understand any differences between the practice of evaluation and what the theory of monitoring and evaluation says about the problems of causation and quantitative versus qualitative methodologies. The responses to these questions also helped me to frame the sample selection and content analyses of evaluation documents by understanding what factors those who write evaluations are focused on in this process and what may or may not be in these evaluations.

3.4.4. Analysing the interviews

After the interviews, I transcribed the audio file into a written transcription. I used this transcription as well as the audio file in forming the initial impressions of the interviews, analysing the way that evaluators said certain phrases or emphasised particular words. I primarily used the written documents to help me code and organise the data. I began by determining the distribution of different words or phrases that evaluators used in talking about evaluation, but this initial approach did not provide sufficiently rich data for a 'theoretically aware' analysis (McDowell, 1998).

I then coded the issues that evaluators raised rather than the distribution of the words and found this to be more useful in building a theory and an argument. Using a 'codes to theory' model (Saldaña, 2010), I developed several main themes about evaluation at the World Bank: how evaluators describe the evaluation processes at the Bank, how evaluators describe themselves and their skills, and how evaluators describe the data they use in determining the ratings. Within these themes, I created subcategories that, for example, differentiated between how evaluators described data collection for project-level and country-level evaluations. By comparing and contrasting these overarching themes and the variations in the subcategories, I identified strong themes as well as inconsistencies in how the roles and processes of evaluation are described by those completing evaluations.

By identifying the differences and similarities in how different evaluators described their role and the process of evaluation, I was able to form more nuanced arguments about the overall role and process of evaluation as well as ways in which these factors may vary. I thought not only about the way evaluators described the evaluation process and their role in the organisation, but also considered the words they used and the anecdotes they used to illustrate their points, which provided clues into the way that evaluators think and what they do. For example, as I discuss in chapter five, many evaluators described their work as an 'art' in describing evaluation as a specialised skill rather than a technical exercise. By understanding what factors those who write evaluations are focused on in this process and what may or may not be in these evaluations, I was able to frame and contextualise the content analyses of evaluations presented in chapters six and seven.

In analysing the interviews, a methodological consideration was the role of the evaluators as potentially misleading sources. Close dialogue depends on those being interviewed being candid and truthful about the status of their roles and their organisation and that "there may be reasons, hidden from researchers, for systematic misrepresentation by individuals and whole groups" (Clark 1998, 79). I had to account for the fact that how the evaluator explains the evaluation process is reflective of how they account for their own actions to an external interviewer (Czarniawska, 1997, 1998). In order to mitigate this possible misinformation in the process of analysing the interviews, I cross-referenced the information that the Bank interviewees provided with the explanatory notes and other resources that the Bank produces as well as the actual evaluations themselves (Clark, 1998).

3.4.5. Elite interviews and positionality

The interviews I carried out are generally classified as elite interviews and, as such, present different methodological issues than other areas of ethnographic or interview-based study (Clark, 1998; McDowell, 1998; Cormode and Hughes, 1999). The previous literature has extensively explored the concept of 'positionality,' which involves considering how one's own identity may shape the research project and its outcome (Rose, 1997). Feminist scholars have often led this discussion by questioning, importantly, how one would know the impact of one's gender on the research process (Schoenberger, 1992; McDowell, 1992). Some scholars have argued that academics hold a position of 'privilege' because they determine the questions and interpret the data (McLafferty, 1995) but, as a student with an 'outsider' status at the Bank, my position was one of generally lesser power and difference (Rose, 1997; Miles and Crush, 1993).

My positionality, or the ways in which perceived aspects of my identity may shape the research process and its outcomes in these interviews was somewhat complicated. There were significant differences between myself, a younger woman, and the evaluators, who were mostly male and several years my senior. These differences in employment status, gender, and age were important for me to consider in how I approached the interviews and how I analysed the data I received in them. Despite the age and gender differences, however, I shared some traits with the interviewees at the Bank: I am from the United States and speak English with an American accent and also, as is true of many of the interviewees, attended a private university and later pursued graduate studies in social science.

To begin with, an important issue in the interview planning was, as McDowell and Schoenberger discussed, the question of "whether to present oneself as expert or an ignoramus" when interviewing elites (McDowell, 1992; Schoenberger, 1992). In order to mitigate this potential inequality of power and knowledge, I often presented myself not as an expert but as an 'informed' interviewer who knew about development and their role but was not a professional evaluator. At the same time, given that the evaluators knew of my status as a student, I encouraged evaluators to explain the details of their role and the process of evaluation to me in detail. In general, Bank evaluators were enthusiastic about demonstrating their superior knowledge about evaluation and the Bank's bureaucracy (McDowell, 1998). I used this enthusiasm as an advantage to elicit extensive explanations from them in the interviews, which proved helpful in illuminating the process to me, especially in helping me understand the perspectives and power structures within the Bank that shape evaluators and their work.

The power differential between me and the interviewees proved to be an important consideration in organising the interviews. Due to my dependence on the evaluators for the research, I did not want to be too demanding in terms of scheduling and access. In their email replies and their comments in the interviews, evaluators expressed that they are conscious of their own importance and often express that they have strenuous demands on their time (Richards, 1996). They often referred to the demanding nature of their schedule and work, saying that they would 'only be able to meet for a short period of time' due to their challenging schedule, although I was able to convince the interviewees to speak with me for at least an hour, or sometimes longer, in the course of the interview (McDowell, 1998). I would have appreciated being able to go with an evaluator to the field or site of a project in a recipient country, likely outside of the United States or Europe, to see first-hand how they collected their data and undertook an evaluation. Unfortunately, evaluators stated in the interviews that it would be impossible to gain access to these trips due to the confidentiality issues involved with the Bank and its clients.

3.5. Content analysis and theoretical considerations

In chapters six and seven, the results of the content analyses of the Bank's evaluations are presented. The results of these analyses explains what these evaluations show about the goals of aid and how success is defined and produced, especially how aid is evaluated at different scales. Content analysis, and specifically discourse analysis, is the main methodology used to analyse the details of the Bank's evaluations in a qualitative way, and to counterpose the findings from the expert interviews with academics and Bank evaluators. Using these techniques, there is a focus in the analysis

on the form of the Bank's evaluations, the specific language they use, and the data they draw on to justify their findings about the impact of aid. In this section, I describe the reasoning for using discourse analysis, detail the selection process for the selection of the documents that I analysed, and finally discuss how I framed this analysis and the way in which I organised the data.

Content analysis is a methodology which investigates data in a specific context in order to draw conclusions about a larger theme or research area (Krippendorf, 2004; Flyvberg, 2006, 2011). This dissertation draws specifically on discourse analysis, which systematically explores the relationship between the narratives that are presented and how these relate to the institutional structures and how these practices are shaped by certain power relations (Locke, 2004; Neuendorf, 2002). Previous studies have used discourse analysis to analyze public communication of organisations and governments. Similarly, in the content analysis for this study, I searched for themes and patterns in public communications, produced by the World Bank. Using discourse analysis helped me categorise and break down the meanings of the 'self-evident truths' that those in power, such as the evaluators in the Bank, use in their work (McGregor, 2009). This methodology was an appropriate tool to help me answer the questions about the production of data within a powerful institution. The advantages of discourse analysis include its ability to reveal the 'unacknowledged' aspects of documents and spoken words, especially of those in power; it can help challenge traditional theory, policy and practice in many contexts (Krippendorf, 2004; Flyvberg, 2006, 2011).

There are some disadvantages to discourse analysis, which were addressed and mitigated through a conscious process. Most importantly, some analysts charge that it is difficult to replicate the findings from this approach because the researcher, who is an

outsider, might not correctly interpret the meanings of different terms (Morgan, 2010). This is an important theoretical concern that is related to the fact that I was an outsider in this research and also that, as others have argued, the Bank's publically available materials such as the ones used in this analysis, adhere to a institutional narrative that is carefully constructed (Mosse, 2005; Broad, 2007). In cross-referencing the findings from the discourse analysis of the written reports with the analysis of the interviews, I moderate this issue by being clear about how evaluators, those inside the process, describe certain terms that are often used in the reports. In addition, the content analysis focused on understanding the framework and the nature of this public narrative rather than taking the information provided in these reports at face value.

3.5.1. Selecting the documents for the sample

In engaging in the discourse analysis, documents were selected from more expansive groups of evaluation documents produced by the Bank. These documents are publically available on the Bank's website in the "Data and Documents" section as well as through links to the Independent Evaluation Group (IEG). In the content analyses in chapters six and seven, I decided to use a selection of documents drawn from a more expansive group of evaluation documents. These documents were chosen as a part of a 'critical' sample, defined as exhibiting many factors that are common to the whole group of evaluations on either the project or country-level (Flyvbjerg, 2006).

In chapter six, a sample of Implementation Completion Report Reviews (ICR Reviews) that professional evaluators complete for each of the Bank's projects are analysed. The ICR Reviews contain the authoritative ratings for the project, the same data that Mosley and other economists have used for their aggregate analysis (Mosley,

1986). I use a selection of recent ICR Reviews in the most common area of project assessment available (in the area of Agriculture and Agribusiness projects). Each ICR Review is approximately ten to twelve pages in length and is presented as a standard form, with numbered sections that the operations staff and later the evaluator, must complete; there are four areas that IEG gives ratings to and identifies in its materials as the most important aspects of a project that reflect its overall success (Carvalho, n.d.; World Bank OED, 2005a).

In chapter seven, the analysis focuses on a sample of the Bank's country-level evaluations in Africa, the region in which the Bank has completed the most country-level evaluations. In that chapter, the analysis focuses on a sample of Bank evaluations undertaken in African countries because Mosley's original micro-macro paradox research used evaluations completed Kenya, an African country, as an example of aid's different impacts at the project and country-level (Mosley, 1986). These country-level evaluations are a review of the progress of the Bank's work in that country over a defined period of time (World Bank OED, 2005b). In the CASCR Reviews, IEG often evaluates the outcomes and other variables along six rating categories from highly satisfactory to highly unsatisfactory, which are derived from international standards (World Bank IEG, Methodology). The CAE and CPE reports are more varied in terms of their format and provide more detailed information about the impact of the Bank's work in a particular country. They describe the World Bank's strategy in the country and the overarching goals that the Bank staff pursued, which are often identified in country-led initiatives such as the poverty reduction strategy papers (PRSP), and, finally, analyse the extent to which the Bank's self-described goals have been met.

3.5.2. Analytical process for the content analysis

Using the qualitative research software NVivo, I investigated both the framework of the Bank's evaluations that are included in the samples as well as the specific language used to describe the impact of the Bank's work at the country-level. I used NVivo both because it is a useful and comprehensive programme and because it is the software package that I have the most experience in using. It is important to note that this comparison is qualitative rather than quantitative: the analysis does not consider the distribution of the ratings, but rather the way in which evaluators justify and codify success and failure in the specific dimensions of the Bank's evaluations.

In analysing the evaluation products that the Bank produces, the theory of monitoring and evaluation helped to explain and contextualise the nature of these evaluations within their field. Overall, the analysis is framed through an investigation of three aspects of project-level and country-level evaluations: the logical framework of these evaluations, their use of indicators, and the way that they describe the issue of attribution or the extent to which the Bank's work can be traced to particular outputs within a recipient community or country. Three aspects of the text are most closely considered: text analysis (how certain concepts are explained), discursive practice (the context in which these concepts are presented) and social practice (how the discourse relates to wider power issues, in this case of the role of evaluation within the Bank) (Fairclough, 2013).

The first step in the analysis that is the basis for chapters six and seven, as in the previous research in the theory of monitoring and evaluation, was to examine the logical framework ('log frame') of evaluations in international development organisations such as the World Bank, which determine what kind of information on which evaluators base

their decisions (Maxwell, 1997; Cracknell, 2000; Grove and Zwi, 2008). I created nodes which captured the common categories on which evaluators evaluated a country and, after determining the steps by which the Bank determines success and the logic of the assessments, I examined the logic and language that is included in these evaluations. Looking within these nodes of the common categories within the evaluations, the analysis dissected how evaluators describe the goals of the evaluation process, how they use indicators, and how they describe the projects and other work that the Bank completed in the country being evaluated. I analysed both what the Bank defines as its impact on the project and country levels and the methods by which the Bank determines its impact, especially which factors the evaluation frameworks prioritise within the institutional context of the Bank. Using these codes, I analysed both what the Bank defines as its impact and the methods by which the Bank determines its impact, especially which factors the evaluation framework prioritises within the institutional context of the Bank. How the Bank's log frame shapes its evaluations by, for example, prioritising quantitative indicators and excluding other considerations of aid impact and beneficiary feedback, is also investigated (Weiss, 1997; Merry, 2011). A key question is the difference between the theory the Bank espouses and the standards by which it evaluates its work. The changing goals that the Bank sets for itself in its client countries are important considerations, as are how they reflect several international agreements: the Poverty Reduction Strategy Paper (PRSP) approach, the rejection of the Washington Consensus, and the acceptance of the Millennium Development Goals (MDGs) and other aid declarations such as the Paris Declaration (Holvoet and Renard, 2007; Conlin and Stirrat, 2008).

Secondly, the analysis focuses on explaining how evaluators 'prove' the results of the Bank's programmes and how it uses of indicators such as the Millennium Development Goals (MDGs) and, from the 'bottom up,' quantitative indicators such as the number of hospitals built or the number of students enrolled (Merry, 2011). At the project level, I used NVivo to analyse these documents to compare how successful and unsuccessful projects are described in the evaluations and how evaluators defend these ratings and the specific words and ideas that evaluators often used to justify certain ratings or to disagree with operations' perspectives. At the country-level, I analysed how evaluators determine the Bank's overall impact within a country rather than just the individual projects and how they describe and contextualise 'top down' and 'bottom up' indicators. At both scales of evaluation, the types of information that is deemed important to use as an indicator are classified.

Finally, the analysis considers the way in which evaluators consider attribution and the link between projects and the observable results on a national level (Davidson, 2000; Weiss, 1997; Cracknell, 2000). The language and the logic that evaluators use is important to explaining how the Bank perceives of this link and what the assumptions are about the ways in which the impacts of the Bank's work may be assessed at and between different scales of aid.

Chapter Four

Evaluation in context:

Deconstructing the organisational role and spaces of evaluation at the World Bank

4.1. Introduction

The World Bank's main evaluation unit, the Independent Evaluation Group ('IEG'), has developed over the past forty years to play an important institutional role in international development, particularly as aid has encountered wide-spread criticism. In the Bank, IEG is responsible for determining and publishing the World Bank's evaluations on all scales (from the project-level to country-level). In this chapter, I analyse the institutional role of evaluation and how it shapes the Bank's evaluation products and ratings. I also deconstruct the 'evaluation space' at the Bank, defined as the role of evaluation practices and where evaluative knowledge is created within the Bank.

Based on the analysis of the interviews with the Bank's evaluators, I identify several institutional forms that shape the production of the Bank's evaluation data. The data created in these evaluations are often used in academic analyses of the concept of 'aid effectiveness,' which has emerged as a standard of what aid must do to 'work.' Drawing on the frameworks provided by institutional analysis (Powell and DiMaggio, 1991; Scott, 2001; Kingston and Caballero, 2009), I assess how IEG's power within the Bank and its regulative role impact the Bank's evaluation products. I begin by considering the history of evaluation within the Bank and the reasons behind IEG's current role. The analysis in the second part of the chapter assesses how the independence of evaluation within the Bank, seen as critically important to the Bank's

accountability, shapes the evaluation space. Next, the analysis focuses on the extent of IEG's regulative role and how this authority and power shapes what information IEG provides. Finally, the challenges that IEG faces within its current evaluation space are investigated, particularly the possible conflicts within the organisation and changes for evaluation in practice. The following chapter in this dissertation provides further insight into the specific evaluation processes at the Bank and the elite evaluation professionals that complete them.

One of the key questions that emerged from the literature is the extent to which IEG is an 'authoritative' figure. I argue that the evaluation space at the Bank is far from neutral and, as institutional theorists argue, the forms of organisations are significant factors rather than neutral arrangements (Powell and DiMaggio, 1991). In analysing the evaluation system at the Bank and how it may be characterised (Leeuw and Furubo, 2008), I argue that IEG's role in the Bank gives IEG a significant amount of power and that this role has important implications for how the data used in 'aid effectiveness' studies is developed in the Bank. IEG's independence and its associated regulative institutional role determine which data is used for evaluations, the authority of evaluation, and the places in which these evaluations are completed. IEG's power over the production of aid effectiveness is not absolute, however, and it appears that the evaluation space is changing as the Bank supports new forms of evaluation and creates institutional spaces outside of IEG that are also responsible for completing and publishing evaluations.

4.2. Methodology for analysing the 'evaluation space'

Two of the main questions in this dissertation are about the context in which evaluation takes place at the World Bank and how the role of evaluation within the Bank influences its evaluation reports. While I will not repeat the full details from the methodology chapter, this section provides an explanation of how the evaluation space at the Bank was analysed for the purposes of this chapter. In order to understand the context of evaluation at the Bank, it was critical to understand both the way that the Bank describes the evaluation function in official documents and how evaluators think about their role and the purpose of their work.

The analysis and the findings here are mainly based on the twenty interviews that were conducted with current and former evaluation professionals at the World Bank and the notes from attending the World Bank's internal Evaluation Week conference in 2013 as an observer. The interviews with the evaluators are expert interviews and the goal was to gain a sense of the 'lived world' of these evaluators in order to understand the role of evaluation from their perspective (Kvale, 1996). In order to understand the nuances of the 'evaluation space,' I asked the questions contained in the 'Data and Measurement' and 'Role of evaluation in the World Bank' sections of the interview questions (the full list of questions is available in Appendix 2 at the end of this dissertation).

In general, evaluators were eager to describe the role and importance that evaluation plays within the Bank and the kinds of information it can provide to the Bank's leadership and clients. The questions that yielded the most interesting responses were often quite simple, such as: "How would you describe the evaluation process? What do you see as the goal and importance of evaluation?" Many evaluators saw this question as self-evident, but it was important to understand, in their own words, what issues the

evaluators highlighted in their responses, as there were subtle differences between how evaluators described this role and how the Bank's materials define it. Given the importance of the evaluation team's 'independence,' which is often emphasised in IEG's published materials (Heider, 2013), analysing how evaluators conceived of the nature of collaboration with other departments was also important. Two important questions for understanding the 'independence' and 'regulatory' role of the Bank were: "Who do you go to for technical assistance if you need it?" and "Who do you collaborate with here at your organization? Do you generally only work with other professionals who do aid evaluation or is there a lot of cross-departmental collaboration?"

The insider's viewpoint that these expert interviews provided was critical to the analysis for this chapter and for chapter five, which follows. In weighing the evidence presented to me in the interviews, however, I had to also take into account that the Bank's evaluators are not neutral sources and may, in fact, be providing misinformation (Clark, 1998). I was an outsider in the interview process and I had to be purposefully conscious of this 'positionality' issue in considering how to pose these interview questions and how to analyse the evaluators' responses (Rose, 1997; McDowell, 1998; Mullings, 1999). As previous works on the role of professionals within development institutions have found (Burawoy et al., 2000; Strathern, 2000; Broad, 2007), there is a certain institutional narrative that professionals often follow in relating their experiences and the value of their work to an outsider. In addition, previous studies have found that development policy is often designed to obscure the dominance of aid organisations who often work around the recipient countries rather than with them (Ferguson, 1994; Escobar, 1995; Harriss, 2002; Mosse, 2005). Given this, it is perhaps not surprising that,

in some ways, the evaluators described their roles differently than the Bank's published explanatory materials do (Mosse, 2005; Harriss, 2002).

The interview questions, therefore, were designed to steer the evaluators away from repeating the official role of IEG as defined by the Bank. The questions focused on eliciting evaluators' opinions rather than a fact-based explanation and the follow up questions often focused on exploring their opinions. In analysing the responses to the questions about the role of the Bank, the goal was to determine not only 'what' evaluators said about the role of evaluation but also 'how' they described it and what these words suggested about the way they see themselves and their professional positions. To supplement and contextualise the interviews, the analysis in this chapter also relies the review of the documents that the Bank uses to describe the evolution of evaluation within the Bank (Grasso et al., 2003) as well as the results of IEG's recent surveys of the World Bank staff that illuminate the views of evaluation of Bank staff working in other areas of the Bank (World Bank IEG, 2011).

4.3. Raison d'être: The evolution of evaluation at the Bank

Previous studies have found that auditing roles often function within organisations to produce legitimacy and that there are specific aspects of the organisational set-up that reinforce this role (Power, 1999, 2003). I was particularly motivated to analyse the World Bank's evaluation space by the prominent research on the micro-macro paradox in economics which uses IEG's aggregated ratings data (based on the percentage of successful ratings from IEG), but does not investigate the context in which these ratings are produced (Mosley, 1986; Subramanian and Ragan, 2008; Doucouliagos and Paldam, 2009). Qualitative studies in Geography and Anthropology

have analysed development organisations such as the Bank (Burawoy, 2000; Broad, 2006; Essex, 2007) and the role of auditing within organisations (Strathern, 2000; Power, 2003), but there are fewer studies on the role of evaluation at the Bank and how the professional practice of evaluation is completed in this specific institutional environment. One exception is Broad's analysis of the Bank's research department in which she argues that the Bank's selection of research papers reinforces a neoliberal development paradigm, led by powerful conventional development economists such as David Dollar (Broad, 2007).

In order to explain the meaning of the data created in the evaluation space at the Bank, it was important to understand the evaluation system at the Bank (Leeuw and Furubo, 2008) and the institutional reasons for IEG's evolution into this particular type of evaluation unit. The World Bank's documents about its evaluation department state that IEG's goals are to be an independent evaluating body, to promote accountability, and to help develop the Bank into a learning institution (World Bank IEG, 2013, 2012, 2011). From these documents, it seems that evaluation at the World Bank developed as an institutional response to external challenges. Public criticism of aid has influenced the institutional form and role of evaluation and the Bank's explanations of the evolution of evaluation within the Bank emphasise this reality. This supports the findings from previous studies, which have found that evaluation teams have had to adapt to particular external pressures and to the increasing attention to evaluation in the field of international aid (Liverani and Lundgren, 2007).

The present-day IEG has evolved since former Bank President Robert McNamara established the Operations Evaluation Department (OED, which later changed its name to IEG). In the early 1970s, McNamara's original plan was to have an evaluation system

that would be a "simple scorecard measuring Bank accomplishments" in order to improve what aid was able to accomplish (Grasso et al., 2003, 1). This 'simple scorecard,' however, quickly developed into more intensive evaluation practices and, by 1975, OED became its own independent unit and developed even more importance within the Bank in the 1990s (Grasso et al., 2003). This shift occurred due to criticism of the Bank's work: many Bank sources cite former president James Wolfensohn's leadership in the 1990s as a major turning point for evaluation at the Bank. In particular, the 1996 Bank meetings in which some declared that "Fifty Years is Enough" was a "shattering experience" for Wolfensohn (Grasso et al., 2003, 73). As a result, Wolfensohn made evaluation a new priority reflected in the five strategic objectives stated by OED that year, beginning with the goal to "[m]ove evaluation to a higher plane" (Grasso, et al., 2003, 65).

Of particular importance to this study, the changing role of evaluation has historically been closely tied to the development of the concept of 'aid effectiveness' which has emerged in numerous treaties as a standard of what aid is expected to provide. As theorists of New Institutionalism argue, in institutions such as the World Bank "interests emerge within particular normative and historical contexts" (Powell and DiMaggio, 1991, 7). The assertion of IEG's independence, and the associated changes in its position and authority, was in response to external criticism that the Bank lacked accountability to its clients. At the same time, the increasing acceptance of a measure of 'aid effectiveness' as a standard for aid programmes influenced the development of IEG's institutional role and goals. The concept of 'aid effectiveness' was influenced by important shifts in the World Bank, especially shifts towards focusing on poverty alleviation and towards seeing the World Bank as a knowledge-producing institution

(Broad, 2007). Aid effectiveness has become critical to the public's understanding of what aid should achieve (Hyden, 2008) and, in the Bank, IEG has become the institutional space through which the Bank produces this information and, importantly, justifies their work to their clients and to the public.

The ways in which the World Bank has decided to change the institutional space of evaluation in order to adapt to criticism and other challenges is important to understanding IEG's role. There are conflicting views within institutional theory on why organisations change. As Kingston and Caballero (2009) argued, institutional theorists have sometimes opposing views as to the reasons institutions change but generally agree that the "impetus for attempts at deliberate institutional design may reflect gradual underlying changes in parameters, beliefs, or knowledge, which result from the spontaneous evolution of the existing institutions over time" (Kingston and Caballero, 2009, 153). The concept of a 'deliberate institutional design' reflects the Bank's evolution towards a comprehensive and independent evaluative space. In the previous forty years, the role of evaluation has been articulated as centrally important not only to 'keeping score,' but, in evaluators' view, to making aid more effective through what is considered to be more comprehensive information.

In the beginning of its institutional history at the Bank, evaluation was located within the operations (reflected in the previous name of "Operations Evaluation Department") in the main World Bank headquarters, allowing for easier feedback between the evaluators and the operations teams. As evaluation became more important to the Bank, however, the evaluations department was given a separate institutional space in order to be more objective in its assessments. This shift in organisational position was

mainly due to increasing pressure to use evaluation as a way to make aid accountable, which the Bank leadership believed they could achieve by making evaluation more independent and thus objective (Grasso et al., 2003). Indeed, official Bank documents conceptualise evaluation not only as an internal exercise but one that is explicitly public and in response to ongoing criticism of the Bank's work. The World Bank's organisation chart reflects IEG's authority within the Bank: IEG reports directly to the Bank's Board of Directors and the Director-General of IEG (currently Caroline Heider) appears at the top section of the Bank's organisation chart (see Figure 4.1).

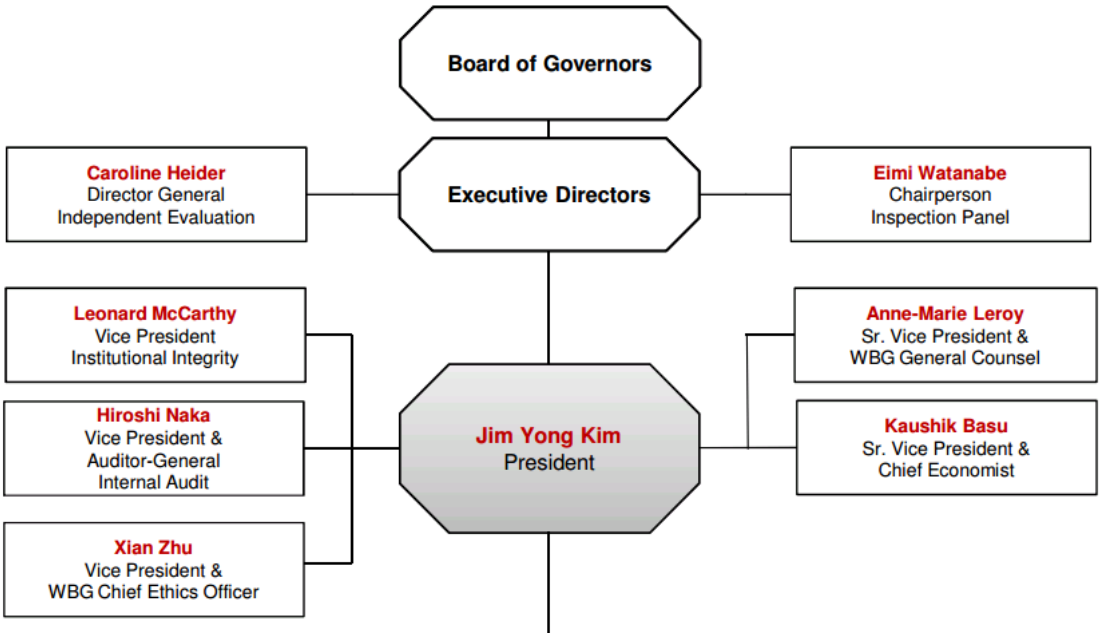


Figure 4.1: The top of the World Bank's organisation chart (World Bank, 2014).

As this chart reflects, IEG's independence gives it an influential position within the Bank near the top of the organisational hierarchy where Caroline Heider, the Director General of IEG, reports directly to the Bank's Executive Directors. This is a significantly higher position within the Bank's organisational hierarchy than the original institutional position of evaluation when it began in the 1970s.

4.4. IEG's independence and the evaluation space

In the present day, the defining characteristic of IEG's space is its independence and separation from the main Bank. In the interviews, I wanted to explore how the IEG's reported independence shapes the evaluation space at the Bank and influences the evaluation data that are produced. As is discussed above, this independence allows IEG to fill an important organisational role: IEG quickly changed from its original set-up in the early 1970s as a monitoring unit embedded in operations to an independent unit as it developed in a context in which the Bank was being publically criticised (Grasso et al., 2003). Within the IEG's documents, IEG leaders emphasise that evaluation must be independent in order to make evaluation more objective. Indeed, in their published materials, IEG leaders often assert that this independence is critical to its ability to do its work, for example a 2011 self-assessment maintained that, "IEG must continuously nurture and draw value from its independence through actions, and CODE/Board must continuously protect that independence" (World Bank IEG, 2011, viii).

This independence is physically defined: the IEG department is housed at 1900 Pennsylvania on the opposite side of 19th Street from the main World Bank building at 1818 H Street in central Washington, DC. The building at 1900 Pennsylvania is the Headquarters of the International Monetary Fund (IMF) and, when visiting an IEG staff member, visitors must announce to the security at reception that they have an appointment with a World Bank staff member in particular as the contact details for non-IMF staff are held in a separate database within the building. Even within its office space, therefore, IEG is separated from other units. IEG and other evaluators did not comment in the interviews on whether this change in the location of their offices made the evaluation functions more or less efficient, since many of them had only joined IEG

years after this change was made. However, nearly all IEG evaluators commented that the physical separation of IEG and operations between different buildings reflected a division in the communications and work of these two functions within the World Bank. As one IEG evaluator noted, "We're in different buildings. That's not by accident. That's very intentional, and it's very much reflective of the culture" (Interview, Current World Bank Evaluator). This 'culture' of independence is consistently reinforced in IEG's written materials and was also characterised as being an accurate assessment of the organisational set-up in the interviews with the evaluators.

One of the initial questions in starting the interviews was whether the IEG's independence was a public relations construct or if it was a situation that World Bank staff actually reported in their explanations of their work within the Bank. In the informal conversations with operations staff and in the question and answer section of presentations at IEG's Evaluation Week, operations staff members shared the view that IEG is independent. For example, several operations staff who attended the conference said that they were not aware of what IEG does and had attended the conference in order to better understand the methodologies of evaluation (Yannias, 2013a). Survey data from when the Bank surveyed 35 Board members, 921 Bank Group staff, and 1,347 external clients in its 2010 client survey reinforces this statement (World Bank IEG, 2011, 5). On the question of IEG's work and results, the widespread perception was that IEG was independent: the survey, only 15% of Bank Group staff said that IEG was not fully independent. As IEG stated, "Similar patterns were found for the four dimensions of independence, and none stand out as a particular problem area" (World Bank IEG, 2011, 17).

Within IEG, most evaluators connected IEG's need for independence to its ability to produce evaluations that are objective. Evaluators generally described this objectivity as the ability to impartially judge a project and to see the project at hand in comparison to other projects. As one current IEG evaluator articulated:

> Thorough task managers, by and large, do an objective job of rating themselves. We have the advantage of a bird's eye view because we have seen a number of projects in different situations. And, we can see it in perspective and we can give it a more objective rating (Interview, Current World Bank Evaluator).

In the interviews, evaluators argued that this objectivity provides the Bank with information that will help it to improve its projects and said that this objectivity helps the Bank to promote its development effectiveness. In this way, IEG fills a critical role for the Bank: IEG is independent in order to provide accountability to the Bank's clients through what are presented as objective assessments. IEG's important role in promoting accountability is reflected in the Bank's requirement that the IEG completes a full evaluation one year after the end of each project, an institutional goal that impacts the nature of evaluation. As IEG's current Director-General, Caroline Heider, emphasised in the inaugural blog post on IEG's blog in September 2013, IEG's model of independence is an important ongoing conversation. As Heider concluded, "In short: structural and intellectual independence is a necessary condition for accountability" (Heider, 2013).

4.4.1. The impact of IEG's independence on evaluations

The independent role that evaluation plays within the Bank influences the information contained in IEG's evaluation reports in tangible ways. As I have discussed, evaluators generally viewed the independence of IEG's role as an important institutional arrangement that helped them do their work, but evaluators' responses in the interviews

indicate four ways in which IEG's institutional role impacts the ratings and other data that are produced. Evaluators' feedback suggests that current project-level evaluation processes: focus on short-term impacts, include data that can be obtained after the project has concluded, limit sharing expertise between evaluation and operations, and make it impossible for evaluators to operate in 'real time' to provide information as needed or to assess progress during the implementation of a project.

First, as a result of evaluation being tied to accountability, evaluators say that there is an emphasis in IEG on the short-term impact of evaluation and they provided examples of projects in which the real impact may only be detectable years after the end of the project. The Bank's requirement that evaluations be completed within one year of the project completion was an ongoing issue that evaluators raised in the Evaluation Week conference as well (Yannias, 2013a). In the interviews, one evaluator gave the example of projects focusing on women's education, where the results were only evident after the women in question had started working several years later, or a clean energy project which had an effect on users' energy consumption only after the evaluation processes ended. Another problem with the focus on short term results identified by evaluators is that the effects of program may be significant in the first year but may then diminish over time. To illustrate this, an evaluator gave the example of a malnutrition program that find that children are very well nourished after one year but that these positive program impacts fade with each year.

Secondly, a consequence of IEG's independent role is that IEG's ratings rely on the data collected by operations teams during the implementation of a project. Several IEG evaluators mentioned that there is often a problem with a lack or quality of data upon which they base their ratings. IEG evaluators only complete ex-post evaluations on

projects and countries: IEG's work only begins after a project is complete, not during the process of development. The timing of IEG's work is viewed by those within IEG and the Bank as being important to protecting their objectivity. Evaluators and Bank materials reflect the view that if IEG entered the development process during the time a project was being implemented rather than only after the project was completed, IEG would not be able to serve as a check and balance to the operations teams in the field. Yet, there are also drawbacks to this institutional requirement. One of the implications of the evaluation only coming at the end of a project that evaluators identified is that the evaluation is not able to consider the many choices and challenges that operations staff may encounter in the course of implementing a project. As one evaluator asked during the Evaluation Week conference, "Should IEG be looking at accountability to meet promises at the beginning of a four year process or at judgments made throughout the process?" (Yannias, 2013a). Another timing-related issue that evaluators raised was that it was often difficult to track the extent to which the agreed-upon project design in the official documents was followed in the field and what the other factors shaping the project might be. As one evaluator articulated:

> I don't think it's as straightforward in many ways as the medical field in that way, but I do think you need a certain sense of the prescription is there. They might follow it and it might not work but at least there's a sense in which they have followed it. And the other thing I think you need to do is follow up. Well, did you take the medicine? Did you take it in the right way over the full length of the treatment and has that made any difference? If not, maybe we need a different diagnosis (Interview, Current World Bank evaluator).

As this quote reflects, there are many questions about project impact that may not be accessible to IEG evaluators due to the separate organisational set up of evaluation. This is specifically due to the fact that IEG's mandate to be independent only allows them to

enter the development process at the end rather than update their appraisal during the implementation.

Thirdly, evaluators reported numerous examples of ways in which the independence of evaluation in the Bank limits what expertise and what project information is included in evaluations. To begin with, IEG's separation means that there is little input from operations staff during the evaluations process other than the information included in the self-assessment and that, if additional information is needed, IEG will contract out rather than ask Bank experts who work in operations. In addition to the physical separation of IEG, many evaluators referred to IEG and operations as having a 'firewall.' They explained this as meaning that there is a separation of information, in particular around the projects and countries currently under evaluation. When asked, for example, who they go to for 'technical assistance,' evaluators often referred to this firewall as a reason that they could not go directly to operations staff. Rather, evaluators report that they would ask within IEG or hire a consultant, perhaps even an external one if there are no available development professionals within the Bank who are experts in the particular field. Evaluators frequently mentioned that they would hire an outside consultant in the case of an evaluation that required technical assistance or expertise beyond their scope:

> Sometimes we do need to go outside and find somebody. As a peer reviewer we can find somebody who knows something about the country, and the topic, but doesn't currently work on it and hasn't worked during the period that we're evaluating. But anybody who was at all involved or has been involved with what we're evaluating we can't draw on, so we have to go outside (Interview, Current World Bank Evaluator).

Evaluators noted that they were more likely to look outside the Bank for experts when this expertise related to the judgment of a particular project, emphasising that keeping

judgments within the scope of IEG rather than in the wider network of the World Bank is key to maintaining IEG's objectivity. While a few evaluators claimed that they would possibly solicit advice from a large range of academics, they often clarified that the academic would be an economist or at least an academic in a related field. As one evaluator explained:

> Well, the language of the World Bank is economics and so an anthropologist who can speak that language can be very successful. … So there's an extra hurdle, I think, for people from other disciplines (Interview, Current World Bank evaluator).

The fact that evaluators think of most of the outside consultants specifically as being economists suggests that the problems they anticipate encountering are not related to understanding the on-the-ground work of the operations staff. Rather, evaluators anticipate possibly requiring assistance with the economic modelling and other technical factors of evaluating a particular project or country.

The firewall between operations and evaluation spaces, however, does not mean that evaluators and operations staffs are forbidden from interacting. While the independence mandate means that evaluators are charged with making their judgment about a project or a country in an independent position, evaluators provided examples of the way in which they interact with other World Bank staff and larger networks of evaluators. In terms of the space of evaluation within the Bank, evaluators emphasise that, while they do not go to operations staff working on the projects they are directly evaluating, they meet with operations staff in informal lunches and in seminars where they may ask general questions of operations staff members. As one IEG evaluator clarified:

> But it's a conservative cautious approach that one takes. You really define quite carefully whatever the question you're asking is. You don't go in with open-ended sort of request. It's for a very specific, targeted issue (Interview, Former World Bank Evaluator).

Evaluators said that the meetings with operations staff were limited to more general development topics or professional development rather than a particular area of the Bank's work.

Finally, evaluators said that as a result of their separation from operations, evaluators often cannot operate in 'real time' during the implementation process or give information that is useful in the case of an emergency. An evaluator gave the example of working in a country where a disease was destroying many of the recipient nation's main export crops and the Bank implemented a program to improve the situation. The agricultural minister called the evaluator because they wanted information right away about the program's effectiveness, but the evaluator said that it would only be possible to complete an official evaluation after the project was complete. As the evaluator recalled, the minister asked, "Are you waiting for people to die and then we start counting the bodies until you get back to us with the evidence?" (Interview, Former World Bank Evaluator). This example reflects that the current institutional set-up of evaluation, in which IEG is central to the Bank's accountability, also limits the information transferred in evaluation products to that data which are captured and relayed to evaluation teams. This also suggests a way in which the methodology of IEG is limited by its independence and the self-reporting nature of evaluation: evaluators are not always able to give real time assessments of projects because they have to delay their evaluative processes until the operations staff is able to supply the data.

4.5. IEG's regulative role

One of the important research questions in this dissertation is about the extent to which IEG exerts an 'authoritative role' within the Bank, if so, what the implications of this role are for the evaluation products. In addition to IEG's independence, one of the main characteristics that define the 'evaluation space' at the Bank is that it is an assumed expertise and its associated power (Linn, 2012). In their descriptions of the ratings process, evaluators emphasised that if IEG did not have the authority to create the ratings, it would undermine its independence and thus its claim to objectivity. Institutional theorists identify three bases of legitimacy within institutions, although they disagree about the most important 'pillar' of institutions: some argue that they are regulative (in that they can establish the rules), normative (in that they define goals and the way to pursue them), or cultural cognitive (in that they define the templates through which meaning is created) (Scott, 2001, 57). I argue that IEG's legitimacy rests on all of these pillars, although the regulative pillar is the most instructive in understanding the Bank's evaluation functions. IEG was established in the context of the Bank's need to be more accountable to its clients and, in filling this function, IEG's ability to "make the rules" -- or in this case, "create the ratings" -- is centrally important to its role within the Bank.

IEG also plays a powerful role in a wider network of evaluation teams within development agencies which augments its strong institutional role within the Bank. The IEG is the largest evaluation unit in the development industry's Evaluation Cooperation Group. In 1996, the Evaluation Cooperation Group (ECG) was established by the heads of evaluation in multilateral development banks (MDBs) in 1996 (Linn, 2012). The Bank's predominance in this network is suggested by the fact that the World Bank creates more research on evaluation than any other aid agency or academic institution

(Yannias, 2013a). IEG is larger than the combined evaluation departments in the African Development Bank, Asian Development Bank, and Inter-American Development Bank (World Bank IEG, 2012).

IEG's authority within the wider development community and regulative role within the Bank is reflected in the fact that there is minimal oversight or review of IEG's work within the Bank. IEG presents its evaluations and ratings directly to the Board and to the public without a review of the data or of the methodologies by which this work is produced. This is a critical aspect of the evaluation space at the Bank: While there is a review of the work of the operations staff within the Bank, there is no such annual internal review on IEG's work, for example by assessing whether their recommendation have proven to be correct. As one IEG evaluator noted:

> I mean that's always, that's something that is a bit of a missing piece, which is worth asking. Who says we are right? Just because we make a recommendation that logically flows from the findings and from hopefully a robust analysis, while we think that that's a good recommendation, we don't actually test whether in fact having implemented the recommendation has led to the expected improvements (Interview, Current World Bank Evaluator).

While evaluators often seemed less knowledgeable about the specific details of the Bank's internal review processes by the Board, a recent blog posting by Heider reports that there is a Committee on Development Effectiveness on the World Bank Group's Executive Board which commissioned an external review of IEG in 2014 (Heider, 2014). She also notes, however, that last such review was undertaken in 2004, so it is an intermittent process rather than a regular review. The main objective of the 2004 review was to provide suggestions and recommendations to the Board of Executive Directors in order to continue to enhance IEG's impact and further strengthen its role as an

independent evaluator of the Bank Group's work rather than to audit specific evaluations (Heider, 2014). Most evaluators did not mention the review of evaluation data as an important issue in their work, but those who mentioned it did not discuss this review process.

IEG's ratings and the Board of Director's high level of attention to these ratings are a visible way in which IEG's authority within the World Bank is enforced: indeed, the ratings are so important to the Bank that promotions of operations staff are based in part on the quality of the ratings for their projects (Yannias, 2013a). During the time in which I engaged in the Evaluation Week conference, it became clear that the operations staff were sometimes resentful of the power these ratings had over their careers, especially when, as a few operations staff noted in question and answer sessions, the operations staff may spend several years managing a project whereas the evaluations staff only spends a few weeks looking at the data (Yannias, 2013a). An operations staff member may challenge a poor rating, but this review is also completed within IEG rather than reviewed by an independent body. Several evaluators noted both that the evaluation process currently creates tensions between evaluations and operations staff, in part due to the possible impact of a poor rating from IEG. As one evaluator explained:

> The task team leader may have been brilliant and have done an amazing job with the support that he or she got. But the institution was just not there to help, and so it comes out with a poor rating. But it gets taken very personally by the task team leader because they are the most wedded to the project. And no manager likes to see a poor project in their portfolio (Interview, Current World Bank evaluator).

Based on evaluators' comments, the fact that IEG's current evaluation process does base its ratings on the quality of the leadership shown by operations or its ability in the field further reinforces the tension between evaluation and operations.

These findings confirm previous research on the theoretical basis of indicators in monitoring and evaluation. The way in which the Bank's ratings are imbued with power is not unique to the Bank and is similar to the way in which indicators function in different development institutions. Using ratings or indicators as modes of governance replaces the role of private knowledge and elite power in decision making with technical, statistical expertise (Merry, 2011; Barry, 2006). IEG's ratings are a review of whether the Bank did an acceptable job (which relies on several factors) while indicators record whether certain clearly defined goals have been reached (for example, how many hospitals were built). As is discussed in chapter six, which focuses on the project-level evaluations, the ratings are often based on multiple kinds of factors, including both the goals reached and how well the Bank's staff and clients contributed to this accomplishment.

The literature indicates that the quantitative nature of the Bank's ratings makes them comparable to many other systems of compliance and evaluation in development organisations and in business (Strathern, 2000; Merry, 2011). In following this type of system, IEG is constructing a certain type of knowledge about development that is similar to the power of an indicator. As Sally Merry (2011, S84) argued: "How indicators are named and who decides what they represent are fundamental to the way an indicator produces knowledge." In the case of the Bank, the evaluators are creating knowledge though a quantitative system that allows the evaluators the power to decide not only what is successful, but also which data are important in determining the success of a project. As I argue in the next few sections, IEG's authority rests on the assumption that evaluators can base their judgment on data rather than on engagement with the places of development or those in the field.

4.5.1. IEG's regulative role and the places of evaluation

Another important way that IEG's regulative role shapes its evaluation products is that evaluation does not have to engage with the places of development because its regulative power is based on its assessment of data. In general, IEG staff does not engage directly with the operations staff or the place of development on the ground and this reflects both the nature of their work as well as how knowledge is viewed within the evaluation space. Geographers have argued that 'place' makes a difference and, as opposed to the way that evaluators describe the knowledge of evaluation, think about development as being a 'situated' knowledge that can only be fully understood within a particular historical context and place (Lawson, 2007). The institutional relationships in the Bank are set up so that IEG is in control of the information: IEG staff complete all levels of evaluations from project evaluations, country evaluators, sector evaluations, and corporate evaluations (of the World Bank as a corporation). IEG staff members perform the majority of project evaluations not from the field but rather from their offices in Washington, DC. Within IEG, different evaluations focus on different types of information and are performed in different spaces geographically and by different staff. IEG is also divided internally between evaluators who primarily work on project-level evaluations and those who work on country-level evaluations. In the interviews, evaluators emphasised that they were less knowledgeable about project-level evaluations if they generally worked on country-level evaluations, for example.

The space of evaluation is, therefore, concentrated in places of power rather than diffused in the places in which development projects take place. In completing project-level evaluations, evaluators obtain their data from the self-evaluations of the World

Bank's operation staff, which is provided on a standard form called an Implementation Completion Report (ICR). This self-evaluation is designed to determine if the project met its self-identified development goals. As one evaluator described the evaluation process:

> I mean, it's difficult to generalise the process, but in essence it's just a desk review of some related literature, to identify the data that allows you to answer the question. Many times it is secondary data; we don't have the time to collect new data (Interview, Current World Bank Evaluator).

It is important to highlight that many evaluators described the project-level evaluation as a 'desk review' or a 'validation.' Using the information provided by operations staff on this form, evaluators make their determinations on the success of projects over the few weeks in which they review the data and write the evaluation in which they validate the ratings of the Bank's operation staff. This Implementation Completion Report (ICR) Review, which takes place in the World Bank headquarters, is the basis upon which the Bank's project ratings are produced and is the most visible evaluation completed at the Bank. From the completed ICR Reviews, IEG selects some evaluations for more thorough evaluation in what is termed Project Performance Assessment Reports (PPARs). Evaluators generally describe the PPAR process as providing better insight into the real impact of a project, but PPARs are only conducted on 25% of projects due to budget and staff constraints (World Bank IEG, 2013; Interview, Current World Bank Evaluator). In contrast to the ICR Review, which is a desk-based review, PPARs include fieldwork and data collecting on the project, often in the country in which the project was completed. Therefore, evaluators only visit the spaces in which development happens on relatively few of the projects that are evaluated within IEG.

There are differences between how evaluators engage with the places of development at different 'scales' of aid within the Bank. While the geography of evaluation activity at the project-level is located primarily in Washington, D.C., it is more diverse geographically in country-level evaluations. Country-level evaluations are a review of the progress of countries' social and economic goals and how this progress relates to the World Bank's projects completed in the country over particular period of time. At the country-level, evaluators sometimes spend time in the client countries, interviewing high level officials. These country-level evaluations can take up to a year of in-country evaluation work, including interviews with high-level officials and NGO leaders throughout the country. It is important to note that, even at the country-level, these evaluations are not focused on the places of development if these places of development are in economically poor communities. Rather, evaluations ask elite within the country -- be they NGO leaders or government ministers -- about their experiences and perceptions of economic progress as a result of the Bank's work. Evaluators who complete country-level reports said that every IEG evaluator involved in the project may not be in the client country for the entire year but may rather rely on the data provided by an in-country staff of both IEG evaluators and outside consultants who are drawn from outside the Bank for country-level evaluations (Interview, Current World Bank Evaluator). In fact, two evaluators who I spoke to who were in the midst of country-level evaluations were available in Washington, D.C., noting that they went to visit the countries they were evaluating on trips consisting of about two weeks at a time. In these two week trips, the evaluators said they mainly conducted interviews with agency representatives and client country government representatives.

The lack of engagement with the spaces of development was also reflected in how evaluators described their work. In the interviews, evaluators usually only reflected on the places of development within the context of explaining the data they used from particular project locations. Based on their responses, evaluators generally perceive of the data they use as being universally applicable and understandable regardless of the context in which it is created or the political context of the deliverables that the operations teams report. This, in many ways, confirms the findings of previous studies that have focused on professionals in international development (Mosse, 2005; Harriss, 2002). Evaluators' perception that certain data reflects the effectiveness of a project even if the evaluator does not spend time on-the-ground reflects the fact that evaluative knowledge at the World Bank has been given a powerful and authoritative role in the organisation. The commonality between evaluation at both the project and at wider levels is that they are focused on spaces of power (i.e., World Bank headquarters in Washington, DC or high level officials in the client countries), not in the actual places where most aid projects take place. This is true in other areas of development as well: in contemporary global governance and aid decisions, an increasing reliance on indicators tends to locate decision making in the global North, where indicators are usually determined (Merry, 2011). The acceptance of this power dynamic within development was reflected in the fact that none of the evaluators I interviewed described country-level evaluations as involving the beneficiaries in the spaces where development projects take place. In the interviews, evaluators described the importance of data and often said that the absence of baseline data was the main obstacle to their being able to do a more thorough evaluation. Due to their institutional role, evaluators are limited in their ability to conceptualise both space and time in the Bank's evaluations because they must do so with the data that has already been collected. Due to time constraints and the wide scope of projects they are

tasked with evaluating, they are not able to collect the data from or interact with the beneficiaries themselves.

Another issue is the nature of the evaluators' expertise, which is primarily related to quantitative evaluation methods rather than particular background knowledge. The evaluators are typically not country experts, although some have expertise in certain areas of development such as, for example, educational development. The global geographical span of evaluation that the Bank completes is possible because of the advances in technology and travel that have accompanied it, making it less necessary for evaluators to be in the places the Bank works to be able to have detailed information about them. The 'time-space compression' noted by Harvey (1989) and other theorists (Massey, 1994) is reflected in the way they describe the timing and global reach of global elites, such as the evaluators in the Bank. These professionals, Massey, Harvey, and others argue, draw additional power from their ability to transverse the spaces of development and business more quickly than most individuals or professionals. Evaluators are one of the "groups who are really in a sense in charge of time-space compression, who really use it and turn it to advantage, whose power and influence it very definitely increases" (Massey, 1994, 149).

As is described in further detail in the next chapter, Bank evaluators describe evaluation as a universal knowledge that can be determined through universal methods. Unlike development geographers (Lawson, 2007), evaluators do not explain development knowledge generally or the information they need for evaluation as a situated knowledge that depends on the context or country in which development happens. Many evaluators mentioned that data, rather than context, should be at the centre of evaluation and few

saw this position as problematic. This reflected a certain mindset in evaluation that development outcomes are quantifiable and that, in the words of several evaluators, that "you can evaluate anything." As one evaluator said:

> Context is very important to people who work in development. They become very interested in their countries as it were ... they think they're unique. It would be like a medical professional becoming obsessed with one patient. They refuse to look at any other patients because, they say, my patient is unique. Which is true, they're a unique person, but it sort of misses the point (Interview, Current World Bank Evaluator).

The 'point' the evaluator refers to in this quote is important to expound upon, as it reflects an important perspective about the role of evaluation within the Bank. According to the evaluators, the main advantage of evaluation is that it is based on 'objective' data, not on an engagement on the ground or on the context in which a particular project was implemented, which they see as being 'subjective' knowledge and less useful in improving aid. As one evaluator claimed, "We have a rich volume of data to review. And that's one of the strengths, I think, of IEG, is it starts with the data management and sources itself" (Interview, Former World Bank Evaluator). Evaluators describe this focus on data as making evaluations more objective and provable rather than based on personal experience or an analysis of the context of the particular project, a knowledge which they describe as being less valid because it does not have an empirical basis. Evaluators describe this data-driven analysis as critical to fulfilling their role within the Bank and the organisational documents about the role of evaluation confirm this perception by focusing on the link between IEG's regulatory power and accountability.

4.5.2. Challenges to IEG's regulative role

The challenges to IEG's regulative role are related to both the relative power of IEG and its evaluators within the organisation and the changing ideas of the role of evaluation as a practice within the Bank. The interviews completed for this dissertation and the Bank's survey results of its staff demonstrate that there are certain challenges to IEG's authority as a knowledge creator and assessor. It is important to note that the research focus is on evaluators' perceptions and on my observations of operations staff at a conference rather than in-depth interviews with either operations staff or Board members. To begin with, IEG's regulative role has created tensions between IEG and the operations staff members, who are judged by the information contained in the evaluations and who are dependent on positive evaluations for their career future within the Bank. In the 2013 Evaluation Week conference, some operations staff were sceptical about the usefulness of evaluation knowledge, with some describing IEG as the 'mean teacher' that only criticised the efforts of operations staff but did not provide any feedback to them that would improve the quality of development or its impact (Yannias, 2013a).

This perception is reflected in survey results among Bank staff. In 2011, IEG conducted a client survey among three groups (Bank Board members, Bank staff, and external stakeholders) to determine the satisfaction and their perspective on IEG's impact (World Bank IEG, 2011). IEG found that Board members have the highest opinion of IEG's impact on the World Bank's effectiveness (with 86% saying the Bank has a moderate, great, or very great impact), while only 65% of staff assess IEG's impact as highly (World Bank IEG, 2011, 22). Echoing what I heard in the IEG conference, World Bank staff was more critical of the 'usefulness' of IEG's work than the Bank's Board

members, with only 20% of World Bank staff saying that they use IEG's work 'very much' or a 'great deal' compared to 53% of Board members who responded this way (IEG, 2011, 23).

The extent to which IEG's recommendations are used to improve or change development approaches was an important issue that evaluators raised in the interviews. In the interviews, evaluators suggested that IEG's recommendations in project reviews may not influence changes in policy design or operations' implementation. There were a range of opinions in terms of how well and how often the Bank used the information gained from these evaluations, with two evaluators saying that this is the research question that may be most unknown to evaluators themselves. As one IEG evaluator expressed, in response to the question of what the Bank does with the information gained in evaluations:

> I think it depends on who you ask. … Some, they say, well, what happens is it gets put on a shelf. … Others, of course, will say that it is absolutely taken into account (Interview, Current World Bank Evaluator).

Evaluators say that country-level evaluations, unlike project-level evaluations, are sometimes discussed at higher levels within the Bank and used to change policies, in part because country evaluations are requested by the clients and in part because the longer time of evaluation and more rigorous approach is often seen as more valid (Interview, Current World Bank Evaluator). In IEG's self-assessment, the worst score IEG gave themselves was in the category of "Incorporating evaluation assessments and findings into recommendation designed to help improve the development effectiveness of World Bank Group programs and activities and their responsiveness to member countries' needs

and concerns." The note reads that "[m]ore can and should be done to encourage learning from IEG's work" (World Bank IEG, 2011, xii).

Perhaps the most critical internal challenge that IEG currently confronts is the rising importance of a new methodology called 'impact evaluation.' This new methodology and the professionals who practice it are bringing other options to the evaluation space at the Bank and decentralising IEG's regulative authority. Theoretically, the increased use of impact evaluation is forcing the question of what the goal of evaluation is: is it an accountability mechanism (as IEG's role is envisioned) or is it a way to identify the linkages between programs and the effects of development? The rise of the use of impact evaluation is, in part, simply because the methodologies of evaluation have changed significantly since the Bank began developing its evaluation department and that evaluation has developed as a profession (Picciotto, 2011). Another reason for impact evaluation's rise in prominence, however, is that the Bank is responding to the ongoing criticism of aid, for example the ongoing criticism stemming from the 'dead aid debate' (Easterly, 2007; Moyo, 2009). Evaluators in IEG were aware of these criticisms of aid and the link between them and the increased attention to impact evaluations, which some view as being more scientific. As one impact evaluator articulated:

> Which is driven by people's impatience with development not working you know this we're not getting anywhere with the MDGs [Millennium Development Goals] and poverty is still with us. Which then leads to scepticism about aid which then leads to pressure for better evaluation and that the research needs to be better. And all of that does have an impact on the kinds of methods we use for evaluation because it puts pressure on evaluation to come up with tangible real time explanations that are useful for decision making as opposed to just for reflection and monitoring evaluation (Interview, World Bank Impact Evaluator).

Impact evaluation uses randomised control trials to determine what would have happened in the absence of the project or programme in order to determine the extent to which there is a causal link between the Bank's work and economic and social change (Banerjee, 2007). Modeled after medical trials for pharmaceutical products, impact evaluation attempts to create a laboratory environment, but in the context of real world development projects. As one evaluator in the impact evaluation teams described, an impact evaluation would design an experiment in two comparable villages in which one village would be provided with a new educational program and the other would not (Interview, Impact Evaluator). Since 2005, impact evaluation has been instrumental in changing evaluation methodologies at the Bank (Banerjee and Duflo, 2011; Banerjee and Duflo, 2009; Banerjee, 2007). The number of impact evaluations completed at the Bank has grown rapidly, increasing fourfold since the Bank began impact evaluations over the course of the past ten years (World Bank DIME, n.d.).

The increased use of impact evaluation has changed the evaluation space at the Bank in two important ways. First, as a result of this methodological change, the evaluation space at the Bank overall is decentralising, or becoming less concentrated within IEG, because significant amounts of evaluation occurs outside of IEG. Although it is involved in some analyses of impact evaluations, IEG is less involved with impact evaluations because it is difficult to resolve IEG's independence with the design of impact evaluations, which embed the evaluation team within the operational team from the beginning. In impact evaluations, evaluators are a key part of the operations staff, working with the staff implementing the project to make course corrections during the implementation and offering suggestions for improvement. Indeed, impact evaluations have a separate institutional space at the Bank for this reason: the Development Impact

Evaluation Initiative (DIME) was created in 2005 by the Bank's Chief Economist Office (World Bank DIME, n.d.) and, in the spring and summer of 2013, this unit was brought into the Bank's main research unit, the Development Research Group. As a result of the different methodology and institutional space of impact evaluation as compared to IEG's evaluations, impact evaluation is also shifting the relationship between evaluation and operations by bringing evaluation into the institutional space led by operations teams.

Secondly, impact evaluation has also changed the use and space of data in evaluation because impact evaluation completes ongoing evaluations during the process of a development project rather than only after a project is completed, as in the case of IEG's evaluation process. A key difference between IEG and impact evaluation teams such as DIME is the relative position of evaluation as they relate to operations work on the ground. This means that, unlike IEG evaluators who are bound by the data that operations teams provide, impact evaluators are able to collect their own data and advise operations teams to make course corrections. Impact evaluation is changing what information is being created by the evaluation process and impact evaluators often describe the information that impact evaluation provides as being more comprehensive than IEG's evaluations in assessing causation and creating knowledge about development impacts. This is, in part, because impact evaluators help to plan and implement the project 'on the ground,' often spending significant time in the community in which the project is implemented and working with operations staff and sometimes in-country agencies (Interview, World Bank Impact Evaluator).

At times, evaluators within the Bank also described impact evaluations as serving a different role from the kinds of evaluations that IEG completes; for example, many

impact evaluators described IEG's work as being an accountability tool. As one impact evaluator described, impact evaluators view IEG as more for the Bank's accountability than as a way to build information about the impact of development:

> Yeah, I mean, [IEG]'s really just an accountability agency, right? To be able to say, 'Yeah, we've checked off the boxes. Yes, it's sort of achieved A-B-C-D,' but it doesn't get into any of the underlying mechanisms, what's happening, why it's happening (Interview, Impact Evaluator).

In describing impact evaluation, impact evaluators often emphasised that impact evaluation provided different knowledge than the kinds of evaluations that IEG completes. In particular, impact evaluators say that impact evaluation has the advantage of 'proving' the conditions in which aid does work to produce real outcomes and identifying the linkages between these outcomes and program implementation. This is changing the nature of what evaluation can do within the context of the Bank and, in particular, those involved in impact evaluations say that, with IEG's methodologies, "there is that missing link between what did that actually produce in terms of what benefit did that have to people" (Interview, Impact Evaluator).

In speaking with IEG evaluators, it became apparent that the question of impact evaluation as the 'best practice' is a contentious one and one which their responses suggest is viewed as a challenge to their authority and their status. As noted above, impact evaluators often say that it provides more rigorous and comprehensive evaluations about what works in evaluation. On the other hand, IEG evaluators describe impact evaluation as a part of a wide range of tools that the Bank uses. As one IEG evaluator said:

> I wouldn't go to the extreme of saying 'it's the only way to measure' or that it's 'best practice.' It's best practice for measuring impact. I wouldn't want evaluation to revolve around impact evaluation (Interview, Current World Bank Evaluator).

Several evaluators in IEG emphasised that impact evaluation is not applicable for every project. Rather, IEG evaluators described impact evaluation as being one of a 'broad church' of methodologies that works for some situations but not for others. As one evaluator illustrated:

> To me, the issue is, impact evaluations -- you know the old saying of the old woman looking for her car keys at night under the streetlight? And you say, 'oh, can I help you?'
>
> 'I lost my car keys.' And you say 'Oh, where did you lose them?' 'Well, one block that way.' 'Well, why are you looking here?' 'Because that's where the light is.' Right? And so, there are some interventions that are really amenable to quantitative measurement … the burden of proof is that you have to show that something is effective. And the only things that you can show are the things that happen to be amenable to impact evaluation design. Then, you're missing out on a whole lot. You're only looking for the keys where the light is, not where you lost them (Interview, Current World Bank Evaluator).

Two evaluators used this 'car keys' analogy in their description of the goal of evaluation in order to convey both the difficulty of evaluation from their perspective and where they think the 'light' in, which is usually related to a data-driven approach. The question of 'where the light is' corresponds to the theoretical consideration of the 'black box' of the link between development and its discernible outcomes and is, indeed, an ongoing question within the theoretical literature (Iverson, 2003; Bourguignon and Sundberg, 2007). In the end, the methodology and the institutional changes at the Bank as a result of impact evaluations are changing the Bank's overall evaluation space that impact not only the kind of information that evaluation provides, but also the institutional role that evaluation plays at the Bank.

4.6. Summary

The Bank's official documents detail how and why evaluation at the Bank is, currently and historically, tied to the production of aid effectiveness (Grasso et al., 2003). The current form of evaluation is based on the logic, expressed by evaluators and by the Bank's materials, that evaluation makes the Bank accountable to its clients. These official documents and the interviews with the Bank's evaluators confirm that there is a particular logic about the goal of evaluation that determines the institutional form of this professional practice in the Bank: In order for IEG to make the Bank accountable, the logic follows, it must be objective and, in order for it to be objective, it must be independent. Following previous studies, I have argued that the role that evaluation plays within the Bank and the way in which it produces legitimacy is important to consider and shapes the practice of evaluation (Powell and DiMaggio, 1991; Power, 2003).

The analysis in this chapter contributes to previous studies on evaluation and on 'aid effectiveness' by highlighting how the role of evaluation within the Bank influences its evaluation products and its production of knowledge about aid effectiveness. Based on the interviews and document review, I found that IEG's independence and its regulative role within the Bank shape the information used to determine aid effectiveness. IEG's independence and its regulative role influence when evaluations are completed, the relationship between evaluation practice and development projects, and the data upon which evaluations are based. This means that aid effectiveness is determined by evaluators who receive data from the places of development rather than having evaluators who are engaged with the places of development in an engaged way. Evaluation decisions are made only after the implementation process and are determined by a

regulative body that not only judges the impacts of aid, but is also given the power to determine the most important factors and the data by which to judge this process.

An important part of the analysis was comparing the explanation of the Bank's role and the implications of this role presented in official documents with the way that evaluators described it. Whereas the Bank's documents describe the role of evaluation in a prescriptive way, the elite interviews provided some insight to the 'inside perspective' of evaluation which illuminated the implications of evaluation filling this particular role within the Bank. In beginning this analysis, I wanted to know not only what the Bank and the theory says about evaluation, but also how their work is shaped by the institutional role that evaluation plays and the mandates of their job within a powerful institution. Evaluators identified these types of examples in their responses to the interview questions and, in this chapter, I have highlighted specific examples of how IEG's current form and role influences the types of data and nature of the evaluations produced by the Bank. In the end, I argue that, when academic analysts (Mosley, 1986) utilise the Bank's ratings and other evaluative knowledge, it is important for them to consider that the ratings are created to fill a particular institutional role and, thus, are, by design, limited reflections of what each project or programme has contributed to the recipient country. The next chapter will focus more specifically on the evaluation process at the Bank and the evaluators' viewpoints on their skills and the goals and importance of their work.

Chapter Five

The art of evaluation:

The evaluation process, evaluation professionals, and the production of aid effectiveness at the World Bank

5.1. Introduction

The previous chapter provided an analysis of the World Bank's 'evaluation space' and how it influences the evaluation reports the Bank produces. There, I established that evaluation, as a practice within the Bank, fills a regulative role and assessed how this authoritative position and IEG's independence in the Bank influenced the nature of evaluation. In this chapter, the analysis shifts to the processes of evaluation and the viewpoints of the evaluators who complete these evaluations. In analysing the role of evaluation in the Bank, I found that evaluators determine the narrative of effectiveness and that, as previous studies have argued, this narrative in supported by an established set of power relations that dictate the narrative and policy decisions of foreign aid (Escobar, 1995; Crush, 1995). As Escobar articulated: "In sum, the system of relations establishes a discursive practice that sets the rules of the game: who can speak, from what points of view, with what authority, and according to what criteria of expertise" (Escobar, 1995, 41).

This chapter is about 'who can speak' about evaluation results at the World Bank and the specific processes by which the Bank's staff communicate to their clients and to the public through official evaluations. The analysis in this chapter focuses on the project-level assessment processes because these are the evaluations which form the

ratings that inform the micro-macro paradox literature (Mosley, 1986). Based upon the interviews with twenty World Bank evaluation professionals and the review of evaluation team documents, I analyse the process of evaluation in the World Bank and how the Bank's evaluation professionals shape the production of aid effectiveness in these evaluation processes. I begin by analysing how evaluators describe the evaluation system the World Bank currently has and the ways in which this system and its procedures produce a certain understanding of aid effectiveness through its formal processes and evaluation products. The analysis then focuses on the evaluators themselves and how these elite professionals understand and potentially shape this process. By analysing how they describe this process, I develop a more comprehensive explanation of the evaluation processes at the Bank that produce the evaluation data. These processes are important to understand because the data are used in academic studies to assess the impact of aid.

5.2. Methodology for articulating the 'insider's view'

In this chapter, I focus on the professionals and the standardised evaluation procedures that shape how the Bank's evaluation teams 'produce' the data about the impact of the Bank's work. The goal is not to criticise the evaluators but rather to explain the ways in which their viewpoints and the evaluation processes they are required to perform produce a particular type of data about the Bank's projects and programmes. It is important for those analysing aid effectiveness and development more generally to understand how these forces shape the Bank's narrative about its own work. The aim is to reshape this debate from one about if aid 'works' (see, for example: Mosley, 1986; Subramanian and Ragan, 2008; Doucouliagos and Paldam, 2009; Picciotto, 2009; Howes et al., 2011) to an exploration of the issues which are considered in determining this data

and the ‘inside view’ of the process that creates it at the Bank. In this approach, the evaluation processes themselves and the decisions made in these processes, far from being neutral, are of central importance. In this section, the specific methodological considerations for analysing the process of evaluation in the Bank are outlined.

The findings are based on interviews conducted with evaluators at the Bank and the observations at the Evaluation Week conference. These inside views were critical to understanding what the evaluation processes measure and how certain elite viewpoints shape this process. The aim was to “gather descriptions of the life-world of the interviewee with respect to interpretation of the meaning of the described phenomena” (Kvale, 1996, 5-6). In the interviews, I was interested in discussing the ‘craft’ of evaluation with evaluators, as opposed to a mechanistic process by which certain boxes are ticked because it is what is required by their employer. In order to be able to comment on the impact of evaluators themselves on the evaluation process, I wanted to understand the nature of the evaluators’ judgments in this process and what informs them: how they view the data upon which they are based and what skills evaluators view as being important. An important issue in this analysis is how evaluators describe the notion of ‘effectiveness’ which is in some ways contradicts the way that previous studies such as Mosley’s (1986) characterise this process.

In order to capture the feedback on evaluators’ views of themselves and their occupation within the Bank, the questions were purposefully broad in the interviews to allow the evaluators to provide examples that illustrated how they view the process and the real-life issues that exist between the complicated world of on-the-ground development and evaluation. In the interviews, I focused on the opinions of the evaluators: why they decided on evaluation as a career, what their opinion was on the

importance of evaluation, and which problems they identified with the current evaluation process at the Bank. The questions that are related to this chapter were mainly from the "Understanding Your Job" and, to some extent, from the "Data and Measurement" sections of the interview questions (Appendix 2). In many ways, these questions served as an 'ice breaker,' because they asked the evaluator to speak about themselves, but the evaluators' responses to these questions were very instructive in illuminating the viewpoints and institutional processes that shape the evaluation data. In these interviews, the evaluators were envisioned not as interviewees per se, but rather as 'shared interpreters' of their work and, in this case, also of the theoretical issues of 'success' in development, for which they are key interpreters. How they described successful projects was instructive in understanding what issues they consider in creating the narration for these concepts and the processes which happen at the Bank.

As in the previous chapter, I had to think about how to weigh the evidence in light of the potential for purposeful misrepresentation (Clark, 1998). It is important to emphasise that, in determining how evaluators see the process, there are many problematic issues of positionality, especially because, regardless of one's occupation, it is difficult to be strictly objective about one's own work. I was aware of the fact that the information I obtained in these expert interviews was shaped by the experts' own points of view, preferences, and how they wanted to portray the evaluation process to an outsider (Audenhove, 2009). I also had to consider how the evaluators, who were mainly male, in powerful positions, and older than I am, would respond to the questions that basically asked them about the details of their work (Rose, 1997). I had some apprehension that the evaluators would exaggerate their importance and minimise the

constraints of the evaluation processes at the Bank in order to maintain their authoritative position in the interview.

This potential constraint was anticipated and planned for in the interview question list and in the interview analysis. In focusing on evaluators' opinions of the evaluation process rather than asking them to only give me a factual account of the approach recommended in the Bank's evaluation department, I gained insight into the process itself and into how these elite professionals understand and potentially shape this process. The way in which evaluators described the processes at the Bank was often similar and provided insight into how the Bank's evaluation team shapes a certain narrative about aid effectiveness as well. As Czarniawska said: "Before a glossy brochure reaches my mailbox, there has been a long discussion about which accounting data to include" (Czarniawska, 1998, 20). With this unseen 'long discussion' in mind, I analysed these interviews in order to understand not only the technical details of evaluators' work but also how evaluators saw themselves in the process. In analysing these interviews, the analysis is grounded in academic literature on the theories of measurement (Bamberger, 2000; Barry, 2006), ethnographies of elite in organisations (Strathern, 2000), and previous research on organisational narratives (Gabriel, 2000; Czarniawska, 1998).

5.3. The project-level evaluation process

In preparing for and analysing the interviews with the Bank's evaluators, I wanted to compare the official narrative of the Bank's report and publically available materials on its evaluation products with how evaluators described these processes (Czarniawska, 1998; Gabriel, 2000). To provide some context to the evaluators' perspective on the evaluation process, I begin here with an overview of the project-level evaluation process

as it is described in the Bank's official documents before delving into the ways that evaluators describe this process. The project-level evaluation process creates the Bank's ratings which are often used in quantitative analysis of aid effectiveness. At the project level, before the World Bank disperses any funding for a project, a Task Team Leader (TTL), a Bank employee who works in the operations team, and other operations staff agree to a proposal with a development objective and the ways in which this objective will be achieved. The development objective is the overarching goal for which the operations team is responsible for achieving and is set forth in the final review of the project. Within six months of the completion of the project, the TTL and other operations staff complete what is called an Implementation Completion Report (ICR) (World Bank, Methodology). The ICR is a self-assessment of the extent to which the operations team has achieved the stated development goal and the data which supports their claim for a successful or unsuccessful result. In an ICR, the operations staff will give themselves a rating on a scale of 1 to 6 which corresponds to the rating of 'highly unsatisfactory' to 'highly satisfactory' and they present the data and any other findings to support their claim in this report (World Bank IEG, 2013).

The next step is where the professional evaluators enter the evaluation process. IEG conducts an independent desk review of all ICRs in a process called the ICR Review. According to the World Bank's documents, this ICR Review process is intended to 'validate' the findings of the self-assessment and to make sure that the self-assessment of the operations staff can be justified with the data provided by the operations team (World Bank IEG, 2013). In these reviews, IEG also gives the operations staff a rating for their project from one to six, ranging from 'highly unsatisfactory' to 'highly satisfactory' based on the extent to which the project fulfilled the stated development

objective and fulfilled the criteria of efficiency and relevance (World Bank OED, 2005a). In validating the ICR, the evaluator does not go into the field to generate their own data or conduct their own assessments of the project but instead bases their assessment on the information and data provided in the ICR. IEG does complete more thorough evaluations on some projects: From the completed ICR Reviews, IEG selects about 25% for more thorough evaluation in what it calls Project Performance Assessment Reports (PPARs). In contrast to the ICR Review, which is a desk-based review, PPARs include fieldwork and data collecting on the project, often in the country in which the project was completed (World Bank IEG, 2013).

5.3.1. Evaluators' explanations of project-level ratings

In analysing the 'window' of evaluation, it is important to recognise what influences what 'we,' or those outside of the Bank's evaluation teams, know about the Bank's projects. Understanding the way in which evaluators describe their occupation indicates that they view their objective review of data as key to their job. In the interviews, evaluators identified several aspects of the evaluation process itself that influence the ratings that have less to do with the impact of a project on a country's economic development and more to do with the evaluation process as it currently stands. As in the previous chapter which considered how the role of evaluation as a function within the Bank shapes the information used in evaluations, these factors were important to identify in the analysis of the interviews as they provide a glimpse into the 'window' of evaluation at the Bank. More specifically, these insights challenge the idea that one can understand aid effectiveness only by considering the Bank's ratings and its calculations of the economic return of a particular project.

First, in their interview responses, the evaluators suggested that the reason for a strong or weak rating on a project may be, to a significant extent, a reflection of the quality of the ICR completed by operations rather than a comprehensive reflection of the impact of the project. Many evaluators indicated that a common concern for them is an absence of baseline data for the project level, which makes it impossible to tell the impact of a particular project. In the absence of baseline data, the Bank's operations staff will complete interviews with beneficiaries about their memories of life before and after a particular project, but evaluators say that the data from these interviews are not as reliable as survey or other kinds of data collected at the beginning of the project. Evaluators report that projects without baseline data or less compelling quantitative data may receive lower ratings, regardless of other positive aspects of a project.

Secondly, it is important to articulate how evaluators explain what they are doing within their role and what they see as the goal and institutional importance of their work. IEG evaluators described the project ratings as 'validations,' which they see as an accountability tool rather than on in-depth evaluation. At the project level, IEG fills the accountability function that many envision for evaluation whereas, at higher levels, IEG is able to engage in more in-depth research. As noted above in section 5.3., evaluators do not assess the project's impact against the needs of the particular community but rather against the original criteria included in the development objective. In the interviews with the Bank's evaluators, most referred to the ICR Review not as an 'evaluation' but rather as a 'validation' and emphasised the prescribed nature of this assessment. As one evaluator described it:

> That's more of an accounting kind of a thing. Look, we had these objectives, how did we reach them? I mean, how well have we reached them? And, how efficiently have we done it? Could someone else have done it more efficiently than us? (Interview, Current World Bank Evaluator)

This feedback suggests that, at least on the project level, the definition of aid effectiveness is limited to "did we do what we promised we would do?" rather than being a reflection of impact or success. In these validations, evaluators at IEG are only charged with figuring out how well a particular project or programme met its stated objectives and was successful. Many evaluators referred to the steps between the programme design and the results of the projects as the 'black box,' especially because they are not involved in the implementation of the project and therefore must rely on the information about the project's planning and its results rather than being able to see the project as it is ongoing.

As was detailed above in section 5.3., the ICR and the PPAR processes are different by design as, due to organisational constraints, a PPAR process cannot be completed for every project. In describing the difference between an ICR and a PPAR, evaluators describe the ICR review process as 'validation' but describe the PPAR evaluation process as 'evaluation.' This distinction is significant as it requires different kinds of data and engagement. The selection process for projects that undergo the PPAR process emphasises the strong role of the evaluation team in determining the validity of the data. Evaluators often say that PPAR selection is due to 'disconnect,' which basically means that there is a significant discrepancy between the operations' staff self-assessment and the IEG assessment in the ICR Review. As one evaluator described the concept of a disconnect:

> In an ICR, they rate themselves, and there's a Harmonized Criteria, a scale of one to six that says, you know, 'fabulous' down to 'sucks.' And they may say, 'fabulous,' and we say, 'sucks,' and that's a big disconnect. … So when that happens, in each ICR Review, there is a box that gets checked on whether it's recommended that this project should have an evaluation done of it, and then a short explanation of why (Interview, Current World Bank Evaluator).

The 'Harmonized Criteria' the evaluator referred to in this case is a set of standardised prescriptions for evaluations that were accepted by the World Bank in 2005 (World Bank OED, 2005a). The details of this 'criteria' are provided in further detail in chapter six but, for our purposes here, it is important to focus on what this quote reflects about this evaluator's view of their own occupation. This example reflects that, in the project-level evaluative process, evaluators have the final say in the determination on the success and merit of the projects.

Thirdly, evaluators' perspectives indicate that the standard against which they are determining the ratings and the impact is generally pre-defined by the outputs anticipated for a project rather than a result of their ability to judge the impact. Evaluators emphasise that, in the current evaluation process, operations staff select which objectives should be met in order for a project to be deemed successful. This means they have some control over how ambitious the project's impacts can be and, according to some evaluators, operations staff sometimes use the system to their advantage in order to ensure a better rating for themselves and their colleagues. The staff that choose the objectives may or may not be the staff that actually complete the project due to the long time frame of the project and the relatively short time-frame of some staff's tenure as Bank employees. Evaluators often suggested that they did not have a mechanism in the current evaluation forms to adequately account for situations in which, while the end result of a project may

not have achieved the development objective, the staff member in charge (TTL) may have done an excellent job of responding to new circumstances and making course corrections.

Evaluators also indicate that, despite their power in this role, they are aware of their limitations in judging the actions of staff on the ground by their position within the Bank. According to the feedback I received from the evaluators, the ratings do not reflect the difficulty of the project and are not weighted higher if there is some progress even in a difficult context. One experienced evaluator used the analogy of a score in a sport such as gymnastics in which an athlete attempting a more difficult skill would get a higher mark even if it is not perfectly executed. In the Bank's current evaluation process, projects are downgraded for not meeting their development objectives even if it is clear that the issue is that the operations team was overly ambitious in the program design rather than that they executed the project poorly. As one evaluator articulated:

> Which is kind of counterintuitive, because we should also be rewarding the project design for having taken such a thorough approach to capturing all of the planned targets. But, as a result, often those projects are the ones that are rated lower because they've been so ambitious in their project design (Interview, Current World Bank Evaluator).

Some evaluators suggested that this reality meant that many operations staff may not take on the more difficult projects or will state a different development goal in order to improve their chances of not being downgraded, which can be a challenge to their career progression.

In the end, a better understanding of the ratings process through the eyes of the professionals that carry out this process indicates that there are many considerations that

go into the final ratings score that may not be related to the quality of the project or, indeed, its contribution to economic growth. I argue that academic researchers and policymakers should be aware of the nature of the data when using the data created through this process to make assessments about the quality and impact of aid (Mosley, 1986). This is especially true because projects are, again, measured against their stated development outcome which is, generally, a specific goal such as "build a road between a town and a city" rather than "increase the economic stability of the town" (Interview, Current World Bank Evaluator). This standard is not surprising and, in many ways, reflects the accepted international framework. The limitations of the evaluation process -- as identified by those who engage in this practice -- raise the question of what one can tell about project's impact from the ratings that the Bank produces and what considerations are not apparent in this rating.

5.4. The country-level evaluation process

Although the country-level evaluations are not involved in producing the Bank's project ratings and are much less visible in the aid-effectiveness literature such as the micro-macro literature that I specifically address in this chapter, here I will briefly discuss the process of the country-level evaluation. I do this to highlight important ways in which it differs from the project-level evaluation process and to show that there are other ways in which the Bank is defining aid-effectiveness in practice. The country-level evaluations differ in methodology, goal, and scope from project-evaluations. To start, as opposed to the process of ICR Reviews which are completed as a matter of course on every project that the World Bank completes, country-level evaluations are completed at client demand or, at times, due to particularly challenging issues that arise in the country.

Whereas the project-level evaluations are based on operations staff members' responses to a common form, evaluators said that the process of country-level evaluations often differed due to the many reasons that country-level evaluations were requested (Interview, Current World Bank Evaluator).

The basic process, however, is that evaluators will be assigned to one country to assess the Bank's overall impact in that country over a defined ten year period. Evaluators generally take about one year to complete a country evaluation and work in collaboration with Bank representatives based in the country. They will begin by reviewing the ICR Reviews for that country in the period of time requested and, if the ICR Reviews indicate that the Bank's projects are concentrated in a certain area, they will focus on a particular area of the Bank's work in that country. After the evaluators review the Bank's data on the projects undertaken in the country, they will collect their own data on certain key indicators (such as, for example, the percentage of women who have access to basic health care). They will also interview key government officials and implementing agencies in the country, which allows them to consider a wider range of issues such as, for example, how sustainable the impacts of the projects were.

According to evaluators, in contrast to project-level evaluations, country evaluations are much more concerned with the impact of the Bank's work than just validating the Bank's own data. The country-level evaluation also uses sector specialists in sectors such as human development, education, and infrastructure to determine the way that specific policies and programmes led to certain outcomes. As one evaluator explained:

> The objectives that you're setting at the country-level are, almost by definition, subject to a much broader array of factors that you can't anticipate than they are on a specific project. So the ability to access what's real, what's realistic at a project level is much, much higher than it is to set realistic objectives than at a political level (Interview, Current World Bank Evaluator).

In discussing the assessment of a country, as this quote suggests, evaluators describe the evaluation process as being far more dynamic and investigatory rather than limited to validating data, as they do on the project level. At the country level, the final evaluation report will consider a wide range of indicators and ways in which the Bank has achieved progress in these areas, drawing on census data, survey data collected in the country, and interviews.

In the interviews with evaluators, evaluators often used terms such as 'impact' or 'unintended consequences' in describing country-level evaluations. In general, they explained that country-level evaluations envision a more involved role for the evaluator than reviewing data that has already been collected by a different staff member. As one evaluator explained about his work writing country-level evaluations:

> Should the question that faces the evaluator be, 'Did the project achieve its intended objectives,' or should the question be, 'What was the impact of that project,' and those are two very different questions. Often learning can come from exploration of unintended consequences. It's much harder to do that, but there's a rich and you know, quite interesting debate about that, and its role in evaluation (Interview, Current World Bank Evaluator).

The evaluators' descriptions of the evaluation process reflect the two different goals for evaluation processes on the project and country-level. While the literature on aid effectiveness often describes this concept as 'success in aid,' I found two different definitions of aid effectiveness to exist in practise. This reflects what Jamie Essex found

in analysing aid effectiveness within the context of an agency, which Essex describes as having dual meanings (Essex, 2010). When describing the Bank's evaluation process at the country-level, and in references to sector-level evaluations, evaluators often articulated a more expansive definition of 'effectiveness,' in part because the evaluation process at the Bank allows evaluators to have more intensive investigations at the country-level.

The evaluators I spoke with were not well-aware of the micro-macro paradox theory and its related publications (Mosley, 1986). However, a few noted that the paradox would not surprise them because of the different nature of what is considered in evaluating a project's success and the issues included in country-level evaluations. Evaluators argued that it is difficult to say that successful projects should 'add up' because the Bank's role in different countries is varied. For example, if the Bank's investment in a country is twenty percent of the overall foreign investment in the country, then it perhaps makes sense to consider if the projects have had an impact on growth. If, however, one is considering a country in which the government has more resources to devote to its own development, than the Bank's role is more minor and, therefore, it would make sense that the projects would not add up (Interview, Current World Bank Evaluator). Overall, evaluators say that there are too many factors included in country-level evaluations that are not accounted for in the final rating in project-level evaluations, such as, for example, the outbreak of a civil war. In addition, the Bank's role in a country may or may not have an economic impact even if the work of the institution was important in achieving other ends.

The Bank's evaluation process in many ways reflects the development of monitoring and evaluation over the years and the literature in this area helped to contextualise the Bank's processes. The theory of monitoring and evaluation is predicated on the claim that evaluation is useful to organisations in order to create knowledge, design better programmes, and, importantly, demonstrate the agency's accountability (Rebien, 1996; Cracknell, 2000). These multiple functions are evident in the way that evaluators describe their roles at the Bank. Some evaluation systems use defined frameworks and indicators that are common to all projects and are generally seen as being more closely tied to the accountability use of evaluation. As George Foulkes said to the Annual Conference of the UK Evaluation Society in 1997: "We now have to define in advance not just what we are seeking to do, but how we shall know whether, and when, we have achieved our goal" (Cracknell, 2000, 53). Beyond this, however, the theoretical basis of monitoring and evaluation theory is the search for causation, explaining the causes and effects of social phenomena and, and evaluation attempts to measure the outcome effect of specific interventions in an increasingly systematic way (Iverson, 2003). This more wide-ranging and in-depth kind of evaluation process is reflected in the Bank's country-level evaluations, but not in its project-level evaluations.

5.5. Elites in the production of aid effectiveness

One of the main issues that I highlighted was the power of the evaluators to shape the evaluation process and the production of aid in the Bank. The interviews with evaluators reflect how evaluation professionals think about their work and suggest specific ways in which they shape the production of the concept of aid effectiveness. In order to better understand the way in which elite professional evaluators shape the production of aid effectiveness, it was important to understand what they saw as the goal

of evaluation and what they viewed as their job as an evaluators. I argue here that evaluators view evaluation as critically important to learning about development and believe that they, as evaluators, have unique skills within the organisation that allow them to understand the impact of aid. I found that evaluators tend to show a strong preference of quantitative data and think about evaluation as being a kind of knowledge that is not necessarily context specific, despite the wide-range of countries in which the projects they evaluate have been implemented.

How evaluators explain the importance of their work was an important narrative to articulate. Evaluators view the their work as more than only reporting if particular projects had a positive economic rate of return, the definition of 'success' in aid effectiveness studies (Mosley, 1986). Rather, evaluators explained that, in their view, evaluation provides the types of knowledge that can improve development in the future. In analysing these issues, organisational storytelling offered an important analytical framework, especially in understanding how wider organisational issues in the Bank are viewed (Gabriel, 2000). In analysing the interviews, I had to account for the fact that how the evaluator explains the evaluation process is reflective of how they account for their own actions to an external interviewer (Czarniawska, 1997, 1998).

5.5.1. Who?: Evaluators' views on themselves

I have argued that the evaluation process emphasises data and operates with two different ideas of what aid effectiveness is and that beneficiaries and the places of development are not often consulted in the evaluation process. Another dimension of the question of IEG's authority within the Bank concerns the way in which evaluators described themselves and the power of their role as professionals. The language that

evaluators use to describe the skills needed to do their job indicates that there is considerable latitude for evaluators to shape the ultimate narrative about effectiveness and that evaluators have the power within the Bank to make this determination.

The evaluators often emphasised the power of their particular role and in describing their job within the Bank and their responses reflected their belief that their skills were both unique and made them well-suited to have the power of judgment. In response to the question about the role of evaluators, for example, one evaluator said, "I'm sure that if you ask this question to people doing evaluation, everyone will have a different answer. I think it's determined by several factors, including personalities" (Interview, Former World Bank evaluator). The fact that an evaluator's 'personality' would influence how they view their occupation within a large organisation suggests that evaluators have the power to shape these determinations. This authority is not unique to the Bank and, although few studies have been completed on evaluators per se, some researchers have found similar structures and power within other development organisations. The role of IEG at the World Bank in many ways fulfils the role of what Barry has termed a "technological zone," specifically a "zone of qualification" that judges the quality of certain standards through technical devices that make it possible for the organisation or government to compare these results (Barry, 2006).

Rather than approaching evaluators as an outsider seeking to understand what the limitations of their role are, I spoke with evaluators and approached the interviews as someone who was interested in entering an 'interpretive community' with them, in order to understand their job from their point of view. I found that evaluators describe themselves as being in a zone of qualification and say this is because they have a set of

wide-ranging skills that allows them to 'know' when a project has been successful. This 'ability to know' reflects the fact that evaluation has now developed into a professional field over the past thirty years, forming organisations such as the American Evaluation Society in the mid-1980s and a growing number of post-graduate study programmes (Chouinard, 2013). It also reflects that IEG relies on its technical expertise in evaluation to justify itself (Barry, 2006). In defining this unique skill set, evaluators often characterised evaluation as an 'art' not a science: In their annual evaluation conference, which I attended in March 2013, the line that "evaluation is an art" was repeated, usually to emphasise that the process itself cannot be fully scientific due to the limitations of data. This phrase was also used to imply that there is a certain skill-set that evaluators have which allows them to make accurate determinations that other staff within the Bank are not in the position to make.

In the interviews, I found that evaluators think of their skill set as unique within the Bank and as not only requiring quantitative skills but also strong abilities to understand people and situations through the interpretation of data. They often describe the practise of evaluation as an art because of the unique combination of these skills. In effect, evaluators perceive that they have an objective eye to understand the development process that other Bank staff cannot have due to their position in the implementation process. In explaining what the 'art of evaluation' is, one evaluator gave the example of how an experienced evaluator was able to deliver a high quality evaluation even in the absence of high quality data due to their skills and insight into evaluation:

> The quality of the evaluation was incredibly good because the person had thought very, very carefully about how they're going to use their two weeks and what the evaluation was for. And they targeted the particular focus sessions that they're

> wanting to do and they had really thoughtful and rigorous approach to how they would be conducted (Interview, World Bank Evaluator).

The fact that evaluators think of themselves as having a 'unique' skill is also true of the words with which evaluators describe their work: Many evaluators and researchers use jargon in describing the evaluation process and one interviewee noted that, upon taking a job at the Bank, he was given a booklet with the "2,000 Most Commonly Used Acronyms" (Interview, Impact Evaluator). As Richard Harper argued in his analysis of a mission at the International Monetary Fund (IMF), it is important for academics to consider the perspective of the professionals who legitimate the data into a public rating because their perspectives shape the end results. Harper noted that, in the IMF, evaluation professionals have the ability to interpret the meaning of the numbers not only to just report them in a straightforward way and that this process indicated their level of authority. Harper wrote: "The process of converting 'raw numbers' into meaningful and 'useable' information constitutes, in part, a *moral transformation* and not just an arithmetical or econometric one" (Harper in Strathern, 2000, 47).

The concept of what Harper refers to as a 'moral transformation' is important to consider, given the evaluators' responses in the interviews. Evaluators justified their notion that they are more knowledgeable about what works in aid than the operations staff by citing their quantitative abilities, which is key to the way that evaluators think about evaluation. Perhaps as a result, a criticism that arose during question and answer sessions at the Evaluation Week conference is that the operations staff claimed that they often know more about the project than the evaluator evaluating it. As one evaluator

explained, there can often be some significant discrepancy between the final determinations of operations and the evaluation team:

> How did you justify that? And, often, it's not justified, and so then we downgrade. And that happens quite often. Statistically, I don't know what the percentage is, but right now I'm working on an annual report. And there's this persistent gap between the ICR rating and the IEG review rating, quite significant. It's, like, at least 10 percent differential. So of maybe 75 percent, let's say, outcome rating satisfactory at the ICR stage, and we would be finding 65 percent (Interview, Current World Bank Evaluator).

In this quote, it is important to highlight the fact that one of the reasons evaluators feel empowered to have an authoritative role is that they are using data and relying on data to justify any claimed impact. In the IEG Evaluation week conference, several operations staff questioned IEG's power to make the final determination. On several occasions, operations staff compared IEG members to the 'mean teacher' rather than as a body that produces the information necessary to improve operations work. This discrepancy in the view of the role of evaluation suggests that IEG evaluators may have a different opinion of their own objectivity and skill set than how their role is envisioned by other staff in the Bank.

5.5.2. How?: Evaluators' influences and experience

In the interviews, I found that there have been some changes both in the nature of evaluation as a career and how evaluators get into the field, with older evaluators often having different training than younger evaluators. In terms of their previous work experience, the Bank's professional evaluators are split between those who came into evaluation from graduate programs in Economics or Public Policy and those who began their work at the Bank in the operations side. In general, fewer evaluators have operations

experience now than in the past: those who began their work on the operations side tend to be older than those who came straight from graduate programs, who are generally in their thirties or early forties. Evaluators who have operations experience referenced this experience as an asset to their work as an evaluator. As one evaluator with more than twenty years of experience in the Bank explained about his previous experience in operations:

> It really helps in understanding the questions that you need to ask because most of what is written is very well written, very articulate because it goes through many, many reviews. And unless you understand what's behind a lot of the phrasing, a lot of the words and what the real life like is actually like it's very hard to come up with a question to which there isn't a almost pre-scripted answer that sounds incredibly persuasive and everything looks fine (Interview, Current World Bank Evaluator).

The more experienced evaluators responded that it used to be that evaluation was what Bank employees did after becoming experienced professionals within the Bank. Evaluation is now a profession that those with high level qualifications are able to do without operational experience, in part because the field is becoming more high profile and more quantitatively driven, with new methodologies, such as impact evaluation, that require specific training.

Evaluators tend to view themselves as 'speaking the right language' and, when asked, evaluators unequivocally identify that language as academic economics. It is difficult to square evaluators' emphasis on the skills of economics with the parallel claim that evaluators make in describing themselves as professionals with a diverse skill set who need a 'broad church' of skills to do their job: evaluators generally describe the goal of their work in terms of quantitative results. I found that evaluators often have a

background in economics as part of their skill set and emphasise that it is the data skills that are essential in evaluation in order to make sure that the results are externally valid. The concept of 'external validity' was repeatedly mentioned in our interviews and many evaluators directed me to the work of economists working at the Bank, such as Martin Ravallion, who specialise in quantitative evaluation methodologies (Ravallion, 2008). Evaluators often express a need to be able to 'prove' what the data states in order to make it as scientifically verifiable and replicable. This is, perhaps, due to the overwhelming influence of economic thought on the Bank's work, an observation that many scholars have made (Woods, 2006). In explaining their motivation for entering the evaluation field, many evaluators note that "I am an economist" and say that they see a parallel between the skill set, particularly the quantitative analytical skills, in their academic field and in their evaluation work.

The fact that the majority of evaluators are economists by training is, perhaps, reflected in the fact that the evaluators often explained their work and their relevant skills in terms of their quantitative abilities rather than their expertise in particular development issues or in the history or politics of specific countries or regions. Only two evaluators I spoke with discussed their expertise in a particular country as being an important factor in their expertise. This reflects their education in a discipline that values quantitative skills over contextual knowledge. As Wendy Carlin and others have argued, economics education has continued to "impose a curriculum that is increasingly remote from what economists now know, and more distant still from the pressing problems that drew our students to economics in the first place" (Carlin, 2013). In keeping with their backgrounds in economics, evaluators prioritise quantitative methodologies, which they describe as being more 'rigorous' and externally valid. As one evaluator explained, one

of the key reasons that having an economics background is helpful in IEG is to be able to understand the newer methodologies such as impact evaluation. As the evaluator stated:

> You look at all fields, in medicine or in other areas. I mean, before they go ahead and do something or advertise a specific type of treatment, there's a lot of testing behind that before learning whether things work or not. So, I think an impact evaluation in other areas, social sciences, of course, including economics and development it's key because basically our resources are limited and we have to get the most out of them (Interview, World Bank Evaluator).

As the above-quote suggests, evaluators often expressed a concern of maximising the outputs of evaluation funds in order to achieve what they see as the goal of evaluation, to make development better through better information. Evaluators generally confirmed that they do believe that they provide more information and knowledge in development through their work and that this is evaluation's contribution to improving development.

In achieving this aim, evaluators' comments reflect that they strongly believe in the integrity of data: While some evaluators say that certain kinds of projects that fall in the categories of 'social capital' or 'social trust' projects may be more difficult to measure, they shared a belief that everything can be measured through quantitative means. This is not to say that qualitative measures are not also used, but these methods are generally less well-regarded when compared to quantitative indicators. As one evaluator's answer reflects, few evaluators at the Bank think about the difficulty of measuring the impact of aid as a key issue they have to consider in evaluation:

> My personal view is that most of these things can be measured with enough ingenuity. A lot of health, for example, was not regarded as measurable only twenty years ago and now we have a whole lot of people doing that, so. There's all this stuff around quality of life, psychological impact... it is possible. Some of the areas that are regarded as particularly difficult which haven't been explored enough yet in development revolve around governance, accountability, but even

> there you've got people creating instruments. I don't know if it's a fundamentally difficult issue (Interview, Current World Bank Evaluator).

Evaluators often claimed that they are able to 'know' about a project's success through the data and to their belief that data is objective and therefore more rigorous proof about the impact of aid. Interestingly, only one evaluator referred to a lack of knowledge about the local situation or country as an issue that might make a project difficult to measure. In general, evaluators expressed a confidence that everything can be measured and, by extension, that there is a certain universality of the ways in which development serves to improve communities and countries which can be measured in a standardised way.

5.5.3. Why?: Evaluators on the goal of evaluation

In the interviews, evaluators often explained that their work is integral to improving the Bank's work. Evaluators expressed that, in their view, the goal of evaluation is to use information to improve aid, reduce poverty, and to improve development effectiveness. Several evaluators found the question of "what you view as the goal of evaluation?" comical, because they viewed the goal of evaluation as very self-evident and also very straightforward. The idea that better evaluation is linked to better development was intrinsic to Evaluation Week. The conference was full of signs that read: "Improving Development Results through Excellence in Evaluation." When asked about the slogan written on these pillars, evaluators agreed with the premise that evaluation would lead to better development results. By way of explanation, they maintained that a major way in which evaluation leads to better development effectiveness is by improving the decision making process about aid and investing in projects that have a higher probability of 'success.' In explaining the importance of

evaluation, one evaluator summed it up as: “lack of evidence or bad evidence can lead to bad decision-making and bad policy-making” (Interview, Current World Bank Evaluator World Bank).

Evaluators stated that the major goal of their work is to make the Bank accountable to its clients and, increasingly, to provide knowledge about what works. Many evaluators expressed a desire that this knowledge be captured in a more rigorous way in order to build a knowledge base about what works in specific contexts. As one evaluator explained:

> What I worry more about is whether the organisation demonstrates that it’s actually learning from what it’s doing, and that’s more of a corporate perspective rather than an individual project perspective. So the accountability that I am more interested in is really the accountability to diligently look at can we learn something from what we’ve done? And then build that into what you’re doing the next time around (Interview, Current World Bank Evaluator).

While some evaluators describe their key goals as improving the knowledge about effectiveness and making the Bank accountable for its work, most go further in their view of what the goals of evaluation should be. Evaluators often explained a direct link between good evaluation and important ends, including: making the outcomes and overall work of development better and making more progress toward alleviating poverty. Evaluators also stated that they thought that improving evaluation methodologies and more comprehensive data could help to improve development outcomes. As one evaluator explained:

> The ultimate goal is to improve development effectiveness. But, you know, two things flow from that. You want the evaluation to be used and you want them to relate to what is going to make the most difference to development effectiveness (Interview, Current World Bank Evaluator).

In order to achieve this 'ultimate goal,' the variables of space or culture do not shape the opinion of the evaluators, who often do not see their work as context-specific. In explaining their 'value add,' or usefulness to the Bank, to an outside observer, evaluators emphasised that they have a unique, strong set of skills that help determine what the Bank has achieved and what lessons may be derived from a particular project or programme that will help future development endeavours.

5.6. Summary

The qualitative research on the production of aid effectiveness has two main findings that contribute to the existing literatures on aid effectiveness and evaluation. To begin with, I found that, in explaining the evaluation process at the Bank, evaluators explain the practice of evaluation as very different in its goals, methodologies, and benefits at different scales in the World Bank. The World Bank's evaluation processes are, in part, a way to provide an accountability measure and, in part, a way to investigate how to make development better, but these different functions are different in practice at different scales of aid. This confirms previous studies: As Essex noted in his research on the Canadian Development Agency, the notion of 'effective aid' is both as a guideline for how to measure aid and how to describe it and, thus, is used to express multiple concepts (Essex 2010, 239).

What is more surprising is that evaluators describe the project-level and country-level evaluation processes in different ways, indicating that these evaluations have different goals. In discussing evaluations at the project level, I found that evaluators view the Bank's ratings as a reflection of not only the impact of the project they are evaluating but also, even more importantly, the extent to which the data demonstrate that the Bank

has fulfilled its stated development objective. This shows that success at the project level is integral to the Bank's need to deliver what it promises to deliver. It does not, however, suggest that the Bank's ratings are intended to reflect the extent to which projects are moving countries forward economically, which contrasts greatly with how previous studies have used these ratings as data. Indeed, it may be that successful ratings, in the aggregate, would not add up to economic growth on a country-level. On the country-level, I found in the interviews that evaluators think there are many different kinds of impacts of aid and that the determination of progress is not only about indicators such as economic growth. Rather, evaluators say that country-level evaluations need to consider many different aspects of development and that it requires a wide-scope of evaluative methods.

Secondly, the interviews showed that the Bank's evaluators view data generation and quantitative analysis as key to both defining aid effectiveness and to being able to prove the results of the Bank's efforts, despite the fact that they view themselves as having a job that requires a diverse set of skills. Evaluators emphasise that evaluation should not only provide accountability but also learning and, more importantly, lead to better development. Overall, understanding the Bank's evaluation system from the perspective of those in power to make these decisions reveals that it is a process controlled by economists who are strongly supportive of quantitative analysis and who believe that their skill set can help to improve aid. It suggests that these professionals are shaping the definition of aid effectiveness to increasingly mean 'what impacts or goals can be proved' quantitatively and that even the learning part of evaluation will move forward using quantitative data and knowledge.

The investigation of these two important influences -- the evaluation process and its elite actors -- sheds light on what shapes what the Bank's evaluations measure and the concept of aid effectiveness as it exists in the Bank. It also opens up avenues for further research on evaluation and the multiple meanings of effectiveness in practice. Despite the importance of evaluation, and the many professional journals devoted to evaluation scholarship, literature reviews have found that "relatively little attention has been given to the theoretical and methodological aspects of aid evaluation" (Rebien, 1996, 5). By analysing how evaluators describe the process of their work, the skills one needs to do their job, and the importance of their work, this analysis contributes to the theoretical and methodological aspects of aid evaluation in the Bank.

Chapter Six

Codifying success in foreign aid projects: A content analysis of the World Bank's project-level evaluations

6.1. Introduction

In the previous two empirical chapters, I analysed the institutional role and processes of evaluation based on the feedback of the Bank's evaluators. In this chapter and the next, the analysis is based on the original evaluation documents from the World Bank at two 'scales' or geographic scopes: the project-level and the country-level. More specifically, I 'disaggregate' the aid debate by analysing how the Bank's project-level ratings are determined and justified in the documents that form the ratings data. The content analysis in this chapter begins at the project-level and centres on a sample of Implementation Completion Report Reviews (ICR Reviews) in the most common area of project assessment available (Agriculture and Agribusiness projects, please see Appendix 3 for a complete list of the projects included in this sample in the content analysis).

The language and logic surrounding these determinations, rather than the distribution of these ratings, are the subjects of the research questions in this chapter. The aim is to analyse and clarify the evaluation processes and frameworks at the Bank, how they function at different scales, and what evaluations measure. This allows for an exploration of what the concepts of 'success' and 'aid effectiveness' mean in practice in the Bank. This is a different approach from the previous research on 'aid effectiveness': the research in mainstream economics on aid effectiveness has quantitatively analysed aggregate data sets of the Bank's reported rate of return for its projects and economic growth in recipient countries, but does not comprehensively explain what 'success'

means in the context of aid projects or if the rate of return for projects is the best measure of such 'success' (Mosley, 1986; Doucouliagos and Paldam, 2009; Subramanian and Ragan, 2008; Picciotto, 2009). The previous research on aid effectiveness, particularly in the micro-macro paradox literature (Mosley, 1986), uses the Bank's self-reported ratings and calculated economic rate of return (ERR) for its projects to indicate that these projects are 'successful.'

There are two main arguments, one about the types of data and information included in the evaluations and the other about the logical framework that the Bank uses. First, I elucidate what the Bank's ratings actually measure and, drawing on the interview responses analysed in the first two empirical chapters, confirm most aspects of how the Bank's evaluators described the evaluation process and data. There are also certain unanticipated aspects of the data that contradict how the evaluators characterised it; these are also examined. Secondly, in investigating the logic of the evaluations, the analysis contributes to the theory of monitoring and evaluation by providing insight into logical framework that the Bank's project-level evaluations use and the complicated way it serves to 'prove' the Bank's impact (Iverson, 2003; Rebien, 1996).

Overall, I argue that the ratings are not a straightforward assessment of the contribution of a project to economic growth. Rather, these evaluations are based upon on a wide range of data and are shaped by particular institutional reporting requirements that influence what these evaluations can tell the reader about a project or programme. Academic researchers and policymakers using the Bank's data to assess 'aid effectiveness' should be aware of the nature of the data and logic used in making these determinations within the Bank.

6.2. Methodology and a review of the data

In this chapter, I analyse 'what' the Bank's ratings mean and then 'how' they are justified in the ICR Reviews. Three aspects of the text are considered in this content analysis: how certain concepts are explained, the context in which these concepts are presented, and how the discourse relates to wider power issues (Fairclough, 2013), in this case, the institutional role and place of evaluation. This comparison is *qualitative* rather than quantitative: this analysis does not assess the distribution of the ratings, but rather the way in which evaluators justify and codify 'success' in the specific dimensions of the Bank's evaluations. While previous research has focused on the reported rate of returns for the Bank's projects (Mosley, 1986), the ICR Review contains significant qualitative information about the decisions made during international development projects and the standards by which these outputs are judged. This information is not captured in a quantitative analysis.

In many ways, the Bank's evaluations reflect the principles of the field of Monitoring and Evaluation and, as such, the logical framework, data, and quantitative nature of the ICR Review are not unexpected (Cracknell, 2000; Rebien, 1996). This analysis contributes to the relevant theory by clarifying how these elements are being developed in practice at the Bank. In order to explain these elements of the Bank's evaluations in-depth, I complete this analysis in three steps. First, I examine the logical framework ('log frame') of ICR Reviews, which determines what kind of information evaluators use to base their decisions (Maxwell, 1997; Grove and Zwi, 2008). By considering the types of data that are used to determine each rating, the meaning of the rating itself and what it measures is clarified. Secondly, I investigate how evaluators 'prove' the results of the Bank's programmes through indicators and other data (Merry,

2011). Understanding the logic of the ratings helps to articulate what, specifically, separates the 'successful' ratings from the others and what this reflects about the meaning of aid effectiveness in practice. Finally, the way in which evaluators consider attribution and the link between the project design and its impacts is assessed (Rebien, 1996; Cracknell, 2000; Iverson, 2003; Bourguignon and Sundberg, 2007).

I was interested in focusing on the context and logic of these reports, not the distribution of the ratings and, in order to analyse the qualitative information in sufficient detail, I restricted the area of analysis to a narrow select. The sample of documents included in the content analysis is drawn from a group of evaluation documents, the Implementation Completion Report (ICR) Review, which determine the project-level ratings. The full versions of ICR Reviews are available to the general public on the Independent Evaluation Group's website: The Bank's Access to Information Policy "allows access to any evaluative material in IEG's possession that is not on a list of exceptions" (World Bank, 2011b, 2) as a part of its accountability priority. The ICR Review is a self-assessment in which operations teams rate their own work and present data that supports the operation staff's claim for its ratings.

In some ways, these documents are uncomplicated to describe. The Bank's ICR Reviews have a uniform layout: they are approximately ten to twelve pages in length and have a predetermined format, in which the operations staff provide and the evaluator reviews the data provided in standard categories using a standard rating system that is consistent between evaluations. A full explanation of each section of the ICR Review is provided in Appendix 4 at the end of this dissertation, but here I will summarise the issues considered in these reports. In these fifteen sections, there are four areas that IEG gives ratings to and identifies in its materials as the most important aspects of a project

that reflect its overall success: Outcome, Risk to Development Outcome, Bank Performance, and Borrower Performance (Carvalho, n.d.; World Bank OED, 2005a). In an ICR, the operations staff rate each rated aspect of the project on a six point scale: Highly satisfactory, satisfactory, moderately satisfactory, moderately unsatisfactory, unsatisfactory, and highly unsatisfactory. In the ICR Review, evaluators in the Independent Evaluation Group (IEG) confirm these ratings or, in some cases, find that the projects should have different ratings than those assigned by operations (World Bank IEG, 2013).

Despite the straightforward presentation, however, there are other factors that complicate the picture of the information that is provided in these evaluations. As I have discussed in the previous two empirical chapters, there are 'unseen' institutional factors that influence the content and logical framework of project-level evaluations significantly. First, in terms of the available data, the ICR Review is a desk review, meaning that evaluators do not engage in on-the-ground work and base their review primarily on the information provided to them in the ICR by the operations staff that are responsible for the project in question (World Bank IEG, 2013). IEG only completes in-country field work on the project-level on the twenty five percent of projects that are selected to be reviewed in a more extensive Project Performance Assessment Report (PPARs) (World Bank IEG, Ratings).

Secondly, in terms of the logical framework by which IEG determines these ratings, IEG's evaluation approach reflects international evaluation principles: the standards provided by the Organisation for Economic Cooperation and Development's (OECD) Development Assistance Committee (DAC), the Evaluation Cooperation Group (ECG), and the United Nations Evaluation Group (UNEG). The table below summarises

the explanation of these key ratings, as described in the Bank's Harmonized Criteria, modelled after international standards (World Bank OED, 2005a).

Table 6.1.: Rating explanation World Bank's Harmonized Criteria		Rating scale
Outcome	Shortcomings in the operation's achievement of its objectives, in its efficiency, or in its relevance	6 Point Scale: Highly Satisfactory (No shortcomings) Highly Unsatisfactory (Severe shortcomings)
Bank Performance	Quality of Entry: shortcomings in identification, preparation, or appraisal Quality of Supervision: shortcomings in the proactive identification of opportunities and resolution of threats	
Borrower Performance	Government Performance: shortcomings in government performance Implementing Agency Performance: shortcomings in implementing agency or agencies' performance	
Risk to Development Outcome	The risk at the time of evaluation that development outcomes (or expected outcomes) will not be maintained (or realized)	Negligible or Low Risk to High Risk (4 point scale)

One of the main aims of this chapter is to clarify what distinguishes positive from negative ratings in the eyes of the Bank's evaluators, based on how they justify the ratings they award in the published evaluations. In each category, in my view, the language describing the difference between positive and negative ratings is vague and does not reflect how the ratings are justified in the Bank's evaluation process. Highly satisfactory project components, for example, have 'no shortcomings' whereas highly unsatisfactory project components have 'severe shortcomings' in each of the components rated. The analysis in this chapter, therefore, helps to articulate what is meant by these standards, what data are considered in determining them, and the methods and logic by which evaluators weigh the data provided to them.

6.2.1. Sample selection

The findings in this chapter are based on the analysis of a sample of the Bank's ICR documents, which I selected from the documents available of the type described above. It is important to explain how I chose these documents rather than others to be included in this sample. I have selected a sample of evaluations using an "information-oriented selection" (Flyvberg, 2011, 307) which is not a random selection, but rather one informed by the analysis of the entire body of evaluation materials available. I have selected an area of ICRs Reviews which are the most common ICR Reviews available online: 649 are for Agriculture and Agribusiness projects while the next most numerous category (Education) has only 450 ICR reviews. The content analysis focuses on the 63 Agriculture and Agribusiness evaluations completed from 2011 to 2013 (Appendix 3), both because I wanted a narrow select of the most recent evaluations and because, as of spring 2014, there was a gap in the evaluations available from 2008 to 2011.

The Bank's filter categorises all of the projects included in this content analysis under the Agricultural group, but these projects include other areas of development outside of Agriculture because each of the projects is broken down by category into a certain percentage of each topic. For example, a project completed in 2013 is categorised in the following way: Central government administration (40%), Sub-national government administration (33%), General agriculture fishing and forestry sector (18%), Tertiary education (7%), Law and justice (2%) (ICR Review 26, section 1). The 18% allotted to 'general agriculture' means that this project is included in the Agriculture category for the Bank's classification purposes, but, as the diverse categories suggest, this project and many others in the content analysis are reflective of a wider scope of activities.

6.3. The meaning of the ratings: data and language

Previous studies have compared the project-level development outcome ratings to economic growth within countries, which assumes a relationship between 'success' (defined as a positive rate of return) and national economic growth but, in the interviews, evaluators reported that there are many considerations in the determinations of 'success' that are not related to its contribution to economic growth (Mosley, 1986). The first step is to clarify which kinds of data are considered in determining what 'success' means in practice. It is important to be specific about the types and categories of data that are used in articulating these ratings, as this sheds light on what these evaluations actually measure about the Bank's projects. The issues the evaluators consider in the ICR Review closely match the requirements of the Bank's evaluation guidelines and the evaluators' feedback, but some of the guidelines are unclear about the differences between 'successful' and 'unsuccessful' projects. In the findings, therefore, I articulate more specifically the types of data that are used in order to make these determinations. The analysis here also reflects a more comprehensive assessment of the entire project evaluation than the data that previous studies in the micro-macro paradox used (Mosley, 1986). As opposed to previous studies that have only calculated the development outcome rating, the analysis in this chapter evaluates all of the ratings provided in the ICR Review (World Bank IEG, Ratings).

6.3.1 Development outcome and ERR

I begin by analysing the steps by which the evaluators must assess within the Bank's evaluation framework and explain the considerations upon which the development outcome rating is based. An important finding of this analysis is that the

majority of the evaluator's considerations concerning the development outcome are not related to the economic rate of return (ERR). While important, the ERR is only one factor among the many that are considered and cited in justifying this rating. Rather, the development outcome rating is more clearly a reflection of the planning, implementation, and results of the project. There are three considerations: the relevance of the objectives and design, the efficacy, and the effectiveness of the project. The data and logical steps by which evaluators describe these determinations are analysed in this section to explain what this overall rating means in practice.

The data by which the development outcome is justified is not unforeseen given the Bank's guidance materials and the way that the Bank's evaluators explained it. In the Bank's logical framework at the project level, the development outcome is based on a combination of data that also includes the project design, outputs, and outcomes and there are three main components that evaluators consider in determining the outcome rating. The development objective is, as evaluators emphasised in the interviews, an accountability tool of the project within the Bank: it is an assessment the overarching goal for which the operations team is responsible for achieving and is the outcome against which the project is evaluated in the final review of the project (World Bank OED, 2005a). The Bank's Harmonized Criteria states that the development outcome reflects "the extent to which the operation's major relevant objectives were achieved, or are expected to be achieved, efficiently" and that shortcomings in this area are related to: that it does not achieve the number or extent of objectives that the Bank leadership promised to the client in the planning stage of the project, that it does not achieve the full economic return, or that it is inconsistent with the country's development priorities

(World Bank OED, 2005a). As Bank economists described in an academic article about the project-level ratings:

> There are a variety of natural concerns about the credibility of these project outcome ratings. A basic concern is that they explicitly measure success in attaining the stated 'development objective' of each project, rather than relative to some common standard across projects and over time (Denizer, Kaufmann, and Kraay, 2013, 8).

In explaining the wide variety of data included in the development outcome rating, I consider the factors, one by one, which are included in the discussion about the development outcome rating in the Bank's standard log frame.

The first issue that factors heavily in the evaluators' consideration of the development outcome is the 'relevance of the objectives and design.' In this assessment, evaluators review the history of the issue in the country and determine whether the project built on previous work in the country, such as the extent to which it continued national policies and followed the Bank's plans and recommendations for the recipient country. In this way, although some evaluators' interview feedback suggested that they do not see context as important, there are ways in which the country-specific issues are included in the evaluations. In justifying positive development outcomes in the written evaluations, evaluators often cite the resources from the client country and explain how the projects within the country are important to particular circumstances in the country. It is important to note, however, that in defining the context and the problems the area is facing, the evaluators often cite official data rather than surveys or interviews with beneficiaries. As one evaluator explains in the descriptive section of an ICR Review, a project which has 'highly' relevant objectives is aligned to the recipient nation's ongoing goals and progress (ICR Review 35, section 3). More specifically, the project follows the government's policy priorities which "have shifted from food self-sufficiency to boosting

farm incomes and redressing the growing income disparity between rural and urban areas" (ICR Review 35, section 3). The project's goals reflect the Bank's FY06-10 Country Partnership Strategy (CPS) and are "consistent with a growing emphasis, worldwide, on the promoting agricultural technology transfer through public-private partnerships" (ICR Review 35, section 3).

Within the assessment of the 'relevance of objective and design,' an importance issue is whether the operation team's self-evaluation (ICR) explains how the Bank's activities would lead to the intended result in a logical way. In assessing the relevance of the project's design, a clear logical linkage between the goals and the activities of the Bank is critical and the evaluator determines the extent to which the evaluation "demonstrates a logical linkage from the activities to be financed by the project to the attainment of the PDOs [project development outcomes], through anticipated outputs and outcomes" (ICR Review 48, section 3). Many evaluators also review the results chain to see if one can logically make the assumption that the Bank's activities would lead to the desired result. As one evaluator concludes of a positive results chain:

> The project objectives were supported by the two main components that could be reasonably expected to promote improved management of protected areas and forests, afforestation and reforestation activities, and conservation and monitoring of carbon sequestration (ICR Review 54, section 3).

As I discuss later in this chapter, therefore, the issue of attribution is considered throughout the ICR Review process, although it is often difficult for evaluators to make a clear and certain connection in practice, as they reported in interviews, often due to a lack of baseline data.

The second important issue in determining the development outcome under the Bank's logical framework is the 'efficacy' of the project. Each ICR lists the outputs and

the outcomes and the extent to which the project met its self-defined goals and the overall impact that the project had on the country or local area. In this, the operations staff must deliver what the Bank has promised and then prove that these objectives were reached through data. In this section, evaluators consider both outputs and outcomes: 'outputs' are the numerical results of the project and 'outcomes' are the wider goals this project contributes to. For example, the 'outputs' of a project would be stated as the number of children who are enrolled in a nutrition program and the 'outcomes' would be the reduction in childhood hunger. The 'outputs' are determined by confirming that targets were reached in several areas in malnutrition and the number of children enrolled in these programs and the 'outcomes' are a decrease in underweight children (ICR Review 52, section 2). The written evaluations reflect that the difference between 'significant' and 'low' efficacy is this: in positive reviews, the evaluator emphasises not only the provable results but also that the operations team proved the results through quantifiable returns. A negative efficacy rating reports, for example, that: "A results framework was prepared but its focus was on outputs and not outcomes. There were no indicators or monitoring and evaluation arrangements to determine achievement of livelihood and income objectives" (ICR Review 36, section 8).

The last component of the development objective rating is the 'efficiency' of the project. In projects deemed to have negative efficiency, evaluators often point to the lack of data, particularly data related to the Economic Rate of Return (ERR). As one evaluator noted, "It is very difficult to conduct a meaningful economic analysis for a project of this nature because of the lack of reliable data on the probability of disease outbreaks occurring or the quantitative efficacy of particular intervention measures" (ICR Review 13, section 5).

While the ERR and the analysis of economic impact is often part of the information that is reflected in the development outcome, there are many different ways in which this data is provided. It is not a standard calculation completed for every project, often due, the evaluator notes, to a lack of baseline data which would allow for a comparison between the original situation and the situation after the Bank's project has been implemented. In reviewing the ICR Reviews in this sample used for the content analysis, I found that many project-level evaluations do not provide the ERR or, if they do, do not have the baseline data to calculate this rate but rather consider the economic impact in other ways. For example, the evaluator may instead analyse a Living Standards Survey to determine the effect of the project's activities on land prices and may also give other data as indications of the efficiency of the project, such as the number of delays and problems with collaborative partners (ICR Review 26, section 5). Overall, therefore, while the ERR calculation is important to the development outcome rating, it is only one of a wider set of considerations that evaluators weigh in deciding this rating. As I have discussed in the analysis of the steps by which the IEG's evaluators explain the rating, it includes a consideration of planning, quantitative outputs, and economic calculations.

6.3.2. Sustainability of project outcomes and evaluation of key actors

In this section, I analyse the types of data and the judgment criteria of the three other categories which are rated in the Bank's project-level evaluations: the Risk to Development Outcome, Bank Performance, and Borrower Performance. In analysing the steps that evaluators take and the language they use in these other ratings, I clarify what these ratings actually measure, what they mean, and why they are important in the overall understanding of aid effectiveness. The development outcome rating is not the only

rating that is important to consider in the overall narrative about aid effectiveness that the Bank produces. In fact, the Bank's main guidance material, its Harmonized Criteria (World Bank OED, 2005a), makes it clear that the ICR Review intends not only to measure the economic return of the project and the extent to which it met its own objectives, but also to assess the sustainability of the project's outcomes and the extent to which those responsible for the project (both the Bank and its clients) took appropriate actions to prepare and implement the project in a way that would achieve the development outcomes.

These other ratings contain important insight into the logical framework and the indicators that are important in the Bank's evaluations. As in the previous section on the development objective rating, I analyse what the choice of particular data reflects about the nature of the evaluation process and the meaning of the ratings. In these other ratings, the wide range of the evaluation criteria that the Bank requires evaluators to consider reflect the contextual and results-oriented issues these reviews are intended to consider. In each category, the evaluator is instructed to consider many issues: technical, financial, economic frameworks, government ownership, institutional support, the relationship with stakeholders, and the monitoring system (World Bank OED, 2005a). These other ratings are measurements of the decisions made in the process, the context in which the project was implemented, and the strength of the monitoring and evaluation systems. In many ways, as with the development outcome, these ratings reflect the recommendations of Monitoring and Evaluation theory and international standards (Cracknell, 2000; World Bank OED, 2005a), but there are also unexpected issues included in the language by which the Bank's evaluators explain their determinations.

To begin with, it is important to explain which issues the evaluators are considering in each of these rating categories. First, the 'risk to the development objective' rating is related to the sustainability of the project and is the extent to which the development outcomes were not or may not be reached. The guidelines articulate that this is rating is intended to measure the probability that the outcomes are sustainable and will not be adversely impacted by the circumstances in the area in which the project is based. In considering the risk to the development outcome, evaluators consider how the environment and difficult circumstances such as an impending civil war may mitigate the outcomes that the evaluators report. Even more importantly, they consider the financial and institutional support for the project and whether it will allow the project to retain its results. In explaining why the risk to a project's outcome is significant, evaluators often focus on financial and governmental issues. For example, one evaluator wrote:

> Though at project completion all dairy park waste treatment facilities built under the project had been at least financed and planned and in some cases implemented, and by the issuance of the ICR in June 2012, all installation was reported as 'near completion,' significant risks remained as to their proper operation and maintenance owing to concerns about adequate compliance enforcement (ICR Review 24, section 7).

In this way, therefore, evaluators consider not only the present condition of the projects outcomes, but its sustainability and support available to continue the project after the Bank exits. This confirms the feedback from evaluators in the interviews on the importance of sustainability in the ratings, although evaluators said that there are often problems with being able to prove this due to the timing of evaluation within the Bank.

Next, the Bank performance rating measures both the 'quality at entry,' which reflects how well the Bank prepared for the operation to achieve the development objectives, and the 'quality of supervision.' The evaluator assesses the extent to which

the Bank staff created a plan for a series of arrangements and contexts -- including technical, financial, and strategic issues -- before entering the field. The Bank performance rating depends on the extent to which the project anticipates challenges and that Bank personnel respond to challenges that arise. In this, the evaluators note the extent to which the operations staff built on previous work and planned the project well and also how well-staffed the operation team was, including the extent of staff turnover and how often the Bank's operation teams made site visits.

Finally, the Borrower performance is assessed, which is an assessment of both the government's and the implementing agency's performance, the quality of the client's preparation, and the extent to which they followed the agreements with the Bank to achieve the development outcomes (World Bank OED, 2005a). In the Borrower performance rating, the evaluator assesses the actions of both the government and the implementing agency. The defining characteristic in this rating for satisfactory borrowers is that they implemented the project in the way recommended by the Bank. The borrowers and implementing agencies are judged on their commitment to the project and the extent to which they are able to form partnerships, seek training when necessary, and monitor the progress of the project in order to change course as needed. As one evaluator summarises:

> The ICR notes that the Government, after some initial reluctance, became fully committed to the project, providing the regulatory framework, appointing a Joint Technical Working Group and establishing administration by a PMU [project management unit] with significant experience in implementing Bank-financed projects (ICR Review 63, section 9).

The Bank and its borrowers, therefore, are evaluated based on how well they planned and prepared for the project in addition to the results that were delivered.

Given all of the factors that are rated in the project-level evaluations, the World Bank is producing a particular definition of success in aid projects that emphasises: attributable outcomes, long-term development goals for the Bank in that area, project planning which connects to country-level development goals, projects which consider the stability of their environment, Bank staff that plans for challenges and responds to them, and in-country clients that are committed to implementing projects in the way the Bank believes is most efficient. In most ways, this is similar to what the evaluators said in the interviews about the importance of evaluation and their views of their jobs as evaluators, but it also clear that there may be some limitations in the Bank's current evaluative framework that the evaluators' feedback in the interviews does not reflect. In this way, there are some contradictions between how evaluators may describe and view their job and the nature and issues that they actually are required to consider in these evaluations.

6.3.3. Sources of the data used to assess the Bank's projects

After determining what the evaluators must do in each of these rating categories, it is important to understand what types and sources of data evaluators draw upon in order to make these determinations. This reflects what issues the ratings actually measure. I find there are two important aspects of the data that evaluators use to justify these ratings overall: first, that the data used to determine the ratings are quantitative data and, secondly, that in using its 'judgment' evaluators tend to focus on the perception of the operations' staff and the quality of the data, to the exclusion of other types of considerations. While the fact that the data is quantitative is not unexpected, the fact that the perceptions of operations staff is often used in the place of data in some ways

contradicts the evaluators' responses and views on the importance of data as expressed in the interviews.

First, in the evaluation documents included in the sample, the Bank's ratings are a reflection of the outputs and aspects of development that can be measured. The focus in these data, across the ratings categories, is on outputs and the deliverable aspects of development. In assessing an agency in a client country, for example, the evaluator offers the following data:

> The project implemented nearly 49,000 infrastructure sub-projects in more than 50,000 villages located in about 4,000 sub-districts in a way that provided temporary employment to nearly 3.8 million local villagers, most of whom (71 percent) were considered poor, despite numerous implementation challenges (ICR Review 28, section 9).

Within the context of these evaluations, these quantifiable outputs are presented without assessing the implications of these deliverables. In stating these numbers, evaluators do not reflect or assess the quality of the deliverables, but rather on the number of units delivered or the number of visits made. The Bank's supervision rating, for example, often only cites the number of visits and the number of technical staff, without considering what these professionals added to the process or what particular ends these supervision visits serve. In stating these facts, evaluators do not explain what the Bank staff completed in these visits or with whom the operations staff met. They also do not trace the impact of particular decisions or the feedback of the beneficiaries.

The ratings are justified in different ways, but the vast majority of the data that are used in the explanations of particular ratings are specifically quantitative data, even for projects that may not have obvious quantitative goals. One can imagine, for example, that the evaluations might contain feedback from interviews with the recipients about the

implications of the Bank's work for their lives. Rather, the ratings are justified with quantifiable ends, for example, the number of new programs the client implemented and the funding level they included for the project in the budget. The table below summarises the kinds of data and the way in which these indicators are presented in the ICR Reviews included in this sample of documents.

Table 6.2: Data used for each rating

Rating	Examples of the types and sources of the data
Development Outcome	• Specific programs and Bank country assessments and how they connect to the project. • List the number of specific outcomes and then how these specific outcomes would contribute to wider goals • Quantitative assessments by the Bank's teams
Risk to Development Outcome	• Government and Bank financial documents • International indicators on corruption and accountability • State operation team perceptions about political stability and conflict • Funding amount that the government has budgeted • Number of functional work sites
Bank Performance	• Length of delays in implementation, if any • Operation staffs' perception of monitoring system • Number of missions • Number of resident specialists • Number of project leaders and staff turnover rate
Borrower Performance	• Number of new government policies and legal frameworks • Funding amount of budgetary commitments • Length of project delays • Operation staff's perception of commitment • Number of workshops and trainings

The table reflects that, in each category, the evaluators focus not only on what was accomplished, but also how often certain actions were completed, and on the specific financial amount of the support. In other words, evaluators create quantitative indicators for goals that may or may not be explicitly quantitatively defined. This suggests how the current framework shapes the evaluation data that authors of academic studies use to assess the impact of aid. Although this limitation is predicted by the quantitative

tendencies and preferences in the theory and practice of Monitoring and Evaluation (Cracknell, 2000; Strathern, 2000), the studies on aid effectiveness do not explain the limitations of the data they use or what it actually measures about the projects which are given ratings.

An aspect of the data that is somewhat less anticipated by the theory of monitoring and evaluation and international standards are the types of data used. Interestingly, the Bank's criteria are unclear in terms of the indicators and which data points are to be used, noting with each category that the evaluator must use their 'judgment' and consider the 'context' of the project (World Bank OED, 2005a). The data that are used in justifying these ratings are in some ways unanticipated because the *perceptions* of the Bank's staff are presented as data. For example, one evaluator stated:

> The ICR argued that the Bank's decision to continue with the project implementation and not close it earlier was risky but justifiable in view of the project objectives, the ownership and commitment demonstrated by the Government, and the lack of an alternative acceptable to the Government (ICR Review 7, section 8).

The use of words such as 'risky' and 'justifiable' suggest that these are the operation staffs' opinions, but these perceptions are not backed up with other data to confirm this assessment. It is important to understand these aspects of the data as these data are the 'facts' which support the ultimate ratings that are presented for each project and included in the data sets about 'success' at the project level in the Bank. The viewpoints of the operations' staff are often stated as though they are a fact, without other facts to confirm that their opinions are, indeed, accurate. The evaluations also omit some important perspectives in their assessment of particular projects, which is not unanticipated by the literature. As some critics of evaluation within development organisations have noted

(Chouinard, 2013), these perceptions are often cited whereas the viewpoints of beneficiaries and the client governments are not.

6.4. The narrative of 'satisfactory' and 'unsatisfactory' ratings

Now that I have clarified the categories of data that are used in each of these categories in the Bank's project-level evaluations and the language with which evaluators explain the data, I analyse how the evaluators describe and justify the reasons for giving particular development outcomes and then the other ratings in the ICR Review to projects. Given the variety of data used and the number of issues that each of these ratings considers, how do evaluators determine the rating and which factors of the many issues described above are given the most weight in this determination?

It is clear from the evaluations that some issues are more important than others in the final determination of each rating. In this section, the 'defining characteristics' that separate positive and negative ratings in each categories are identified. By 'defining characteristics,' I mean the issues that evaluators discuss in their justification of awarding 'satisfactory' and 'unsatisfactory' ratings in each rating. Following the order in which it appears in the evaluations (development outcome, risk to development outcome, Bank performance, and Borrower performance), the issues that distinguish positive ratings from negative ratings are explained.

This analysis is, in my view, necessary due to the vague language in the Bank's guidelines that accompany its logical framework. As is reflected in table 6.1 above, the Bank's evaluation criteria provides some guidance in terms of the issues that the data should reflect. It is, however, less instructive in terms of what separates 'satisfactory' and

'unsatisfactory' projects, noting only that various levels of 'shortcomings' direct the evaluator to give a particular result. In terms of the specific steps of the decision-making, the criteria state that:

> The evaluator must use judgment in weighing possible shortcomings in the achievement of the operation's objectives, in its efficiency, or in its relevance, and arrive at an assessment of how they affect the overall rating (World Bank OED, 2005a, 1).

It is important to understand what particular issues are provided as evidence that 'tips the balance' to justifying positive and negative ratings in each category, given the variety of data the evaluators consider.

Overall, in the Bank's written project-level evaluations, the issues that are most important across the ratings are related to showing through data the deliverable outcomes of the project and the planning and implementation of this project. For each rating, the outcomes and the degree to which the operations staff is able to monitor and quantify them are a critical aspect of each of these rating categories and the defining factor that often separates the 'satisfactory' and 'unsatisfactory' outcomes. In explaining a satisfactory development outcome, evaluators generally stated that the goals have been met and cite the evidence that demonstrates how they have been met. As one evaluator summarised in the written evaluation of a successful project:

> The project's objectives were consistent with Bank and Borrower commitments to reduce rural poverty and the results chain made allowance for the actions needed if the incomes and productivity of project beneficiaries were to rise. The income objective was fully met and the productivity objective was substantially met. Based on the rate of economic return and the evidence of cost effectiveness (cost per beneficiary, M&E cost) efficiency was rated substantial (ICR Review 16, section 6).

The evaluation documents included in the sample show that the operations staff's ability to demonstrate the project's outputs and to connect these outputs to the Bank's work is what separates projects with outputs deemed satisfactory from those deemed unsatisfactory. This confirms what I often heard in the interviews: IEG evaluators emphasised that the ratings are a product of meeting the stated development goal and the development outcome rating is most reflective of the extent to which the operations team did what it promised.

Another defining characteristic which distinguished 'satisfactory' from 'unsatisfactory' development outcome ratings is often the effectiveness of the project's design and the quality of their monitoring and evaluation systems. In criticising the development outcome, the main issue the evaluators identify is that the operations staff did not complete a systemic review and failed to include convincing data to support their claims in the ICR. In the written evaluations, a common reason that evaluators give to justify their downgrading[1] of a particular development objective is the lack of evidence to support the operation team's assessment of the objectives reached or their economic impact: Many project activities were not completed, and there is little evidence of project efficacy. The economic analysis was not meaningful and there were inefficiencies in implementation (ICR Review 19, section 12). Often the issue is neither the outcomes nor the stated perceptions of the results of a particular project, but rather the data and the level of attribution that can be proved. For example: "Based on the evidence provided in the ICR, there is considerable uncertainty on the extent to which the outcomes related to the second objective can be attributed to the project" (ICR Review 58, section 12). The

[1] In the ICR Review, evaluators have the option to 'downgrade' a particular rating or disagree with the rating provided by the operations staff if the evaluator finds that the higher rating is undeserved based on the data presented in the ICR Report.

‘uncertainty’ about attribution in this case is the main reason for the lower rating, rather than a proven failure of deliverables for this project.

It is important to compare and contrast the narrative of the other ratings provided in the ICR Review as well. How evaluators assess the risk to this development outcome, for example, is important to consider, as it speaks to the sustainability of the project. Projects with development outcome ratings of ‘moderate’ or ‘low risk’ differ in two key aspects from projects seen as posing higher risks to development. In giving a positive rating to a project, evaluators do not mention environmental issues that may be problematic and evaluators find that there seems to be a strategy for continuing the institutional support and funding for these projects after the Bank completes its engagement. As an evaluator writes about one project deemed to have lower risk:

> This system is supported by the important political parties and the government. There is continued support from the State budget. Furthermore, the growth in Bank linkages and assets is continuing, and more livelihood-enhancing initiatives are starting and increasing in scale (ICR Review 14, section 7).

In explaining projects with high risks to the development outcome, on the other hand, evaluators stress the environment and the problematic nature of the kind of project that has been implemented. In justifying this rating, evaluators also point to the unpredictable outcomes of the work, either due to the type of work or because the outcome relies on other circumstances which may or may not happen. Many evaluators charge that the operations staff did not provide an adequate risk assessment of their development outcomes in the ICR, especially in terms of the long-term stability of the political situation in the host country. The issue is often not only that the operations team has not implemented the project well or that the government is not supportive, but that there is a high chance that contextual concerns, such as an ongoing civil war or tensions, threaten

the development outcomes. Another potential risk to the stated outcome that evaluators note is that there are inherent risks in the kind of work in which the project is engaged such as, for example, in farming. As one evaluator noted, "Inadequate maintenance of dirt roads is a problem, and unless the maintenance problem is resolved, development outcome is significantly at risk" (ICR Review 38, section 7).

Next, in justifying the Bank performance rating, the evaluators determine the Bank staff's ability to draw on previous development lessons, respond to challenges, and enable a strong monitoring system to separate satisfactory from unsatisfactory ratings in this category. Projects in which the Bank's quality of entry was deemed 'unsatisfactory' differ from the 'satisfactory' projects in several ways. Many 'unsatisfactory' ratings in this rating area are for projects in which the Bank acted in an emergency situation in which they did not consider the relevant permits they would need or the institutional challenges to the project. Evaluators write that these projects also suffer from a weak Monitoring and Evaluation system, which is meant to have been constructed during the planning phase, and suffer from issues such as insufficient indicators.

In assessing the borrower, the key issue that evaluators often mention is whether or not the government and implementing agency have taken concrete steps towards supporting the project. The governments with the highest ratings show their commitment not only by following the mandates of the project, but by taking political action through policy formulation and changing existing programs and bringing in new resources to address needs as they arise. In IEG's estimation, the most 'successful' implementing agencies combine expertise and dedication. The ICR Reviews with these ratings use words such as 'competent leadership' and emphasise the 'strong commitment' of these agencies. In some cases, although the implementing agency ultimately ended up with a

satisfactory rating, they had initial hurdles created by a lack of experience with the Bank's organisational requirements but made up for this as a result of training over time which, evaluators note, is provided by the Bank. As one evaluator summarised in describing a positive rating:

> Government commitment was evident during project preparation and implementation, with high-level, multi agency meetings and workshops to set the project's scope and shape its concept. Before the project started, the Government has set up the National Steering Committee on Avian Influenza to provide oversight of the project and serve as a technical advisory board and had developed a National Plan on avian influenza (ICR Review 7, section 9).

In evaluating projects with unsatisfactory implementing agencies, conversely, evaluators cite problems with the comment or the competence of the implementing agency.

As I have noted, many of the findings in the above analysis about the individual ratings are consistent with the Bank's guidelines and the evaluators' feedback in the interviews. As evaluators reported in the interviews, they are looking for data to support the fact that the Bank has fulfilled its goals, has done so in a sustainable way, and that the Bank and its clients have been proactive in fulfilling the project goals. These standards are apparent also in the language and logic by which the ratings are defended and justified in the written evaluations.

Yet, there are some surprising ways in which evaluators explain and justify their decisions within the ICR Review which contradict the measured and data-driven way that the evaluators described these determinations in the interviews. An important aspect of the way in which the ratings are determined is related to how evaluators use the Harmonized Criteria in their explanations to distinguish between positive and negative results. The Harmonized Criteria is often used to justify why particular ratings must be given for a particular project, but the data which are the basis of these determinations is

often based upon the operation team's perceptions rather than economic data. For example, as one evaluator explained:

> The relevance of objectives was substantial. Yet, the modest relevance of design and modest efficacy of the original PDO [project development objective] were explained by the weak causal link between these outputs and its ambitious expected outcomes, as well as the limited achievements for most project activities. There were also shortcomings in efficiency, which was also rated as modest. Based on these ratings and the IEG/OPCS Harmonized Evaluation Criteria, IEG rates the overall outcome moderately unsatisfactory (ICR Review 39, section 6).

As this quote reflects, evaluators often weigh the ratings according to a pre-determined formulae rather than considering the overall impact of the project. The fact that the evaluators often use the language and framework provided by the Bank is not surprising (Holvoet and Renard, 2007), but the data by which they support these assessments and provide judgments is often unexpected in the sense that it is based on the evaluators' subjective perception rather than objective data.

This is important to recognise in the overall narrative and in assessing the resulting ratings: that the end result of the rating may be due to a mathematical weighting rather than an assessment of the impact of the project, its sustainability, or the performance of its key actors. In the ICR Reviews, rather than considering the contribution of the project to country-level goals or even the strength of the project based on the elements they reviewed, evaluators often justify their project rating by re-stating the three sections within the development objective section and weighing them in terms of their 'success rate.' This suggests a possible limitation of the evaluation process as it currently stands: the ratings are often justified not with an assessment of the data, but rather by restating and discussing the viewpoints of the operations staff.

6.4.1. Attribution in the logical framework

The final issue that I consider in terms of the logic of the ICR Reviews is how the Bank's evaluations develop the concept of attribution in practice, which is of critical importance in the theory of monitoring and evaluation (Bourguignon and Sundberg, 2007). This issue is further developed in the next chapter, which focuses on country-level evaluations and in the analysis on the question of the extent to which successful projects 'add up' to measureable success on the country-level. In the interviews with the evaluators and in the written evaluations, it is clear that attribution is a critical issue that evaluators consider and yet, at the same time, the ICR Reviews reflect the difficulty of proving attribution and suggest a difference between the theory and practice of evaluation. In the ICR Review, evaluators consider the strength of the monitoring and evaluation system in every rating category and, for a positive development objective to be reported, there needs to be a strong monitoring framework with a clear causal chain that states the outcomes that are to be achieved and clarifies the link between these outcomes and the projects that the Bank funds. In considering the relevance of a project objective, evaluators often refer to the link between indicators and the outcome of the project. For example: "The indicators were focused more on actions rather than outcomes: e.g. there was no indicator for increases of fiscal revenues; or, the indicators for improvement in the business environment were exceedingly timid" (ICR Review 21, section 3).

Despite the attention to attribution in each area of the ICR Review, in the written evaluations, evaluators identify problems with the available data upon which they must assess a project. This confirmed a key limitation with the evaluation process at the Bank that the evaluators repeatedly mentioned in the interviews. There are several obstacles to

proving attribution which are reflected in the ICR Reviews in the content analysis. First, as evaluators mentioned in the interviews, there are often other initiatives from the client states and other organisations that have projects that occur simultaneously. In several evaluations, evaluators note that it is difficult to ascribe the outcomes reported to the Bank's project as opposed to the other initiatives. As one evaluator wrote about a project:

> Not all achievements reported below can be attributed to the Global Environmental Facility [GEF] project because pollution control measures were supported also by two other Bank projects: the Tianjin Urban Environment and Development Project and the Water Conservation Project. These projects ran in parallel and supported investments in wastewater treatment plants in the same locations (ICR Review 31, section 4).

Secondly, in the interviews, several evaluators referred to the steps between the programme design and the results of the projects as the 'black box,' especially because they are not involved in the implementation of the project and therefore must rely on the information about the project's planning and its results rather than being able to see the project as it is ongoing. As in the interviews, evaluators' written work often notes that it is generally not possible for them to prove the 'counterfactual,' or what would have happened in the absence of a project, with the data they have access to (Banerjee, 2007). As one evaluator wrote: "It is difficult to establish a counterfactual -- to establish whether the situation would have been worse without the operation -- since all related outputs were achieved at or after closing" (ICR Review 39, section 4).

In sum, the way in which attribution is articulated in these evaluations reflects that it is difficult to prove in the 'real world,' both because of data problems and because of the realities of multiple actors within the same space. This suggests an important difference between the theory and practice of monitoring and evaluation, in which attribution is an important goal (Rebien, 1996). As will be explained in further detail in

the next chapter which compares project- and country-level evaluations, the language of the Bank's evaluations reflect that the process of proving attribution is even more difficult at the country-level. This finding also poses a challenge to the assumptions and the analytic processes of the micro-macro paradox research (Mosley, 1986).

6.5. Summary

In this chapter, I focused on two main arguments: one about the contents of information in the ICR Reviews and the other about the logical framework that the Bank uses in these validations. In comparing the actual data contained in these documents with the interviews with evaluators and official Bank guidance materials, I confirmed the evaluators' feedback that the ICR Review is an accountability tool and that it represents a 'validation' rather than an in-depth evaluation. In contrast to the approach of previous aid effectiveness studies that only focus on the distribution of projects with positive ERR as an indication of the overall impact of the Bank's projects, I found that, at the project-level, the Bank's ratings draw on a wide scope of data in defining and determining its ratings. The analysis of the Bank's ratings suggests that 'successful' project ratings reflect a wider set of considerations: the planning and preparation of projects, the quantifiable benefits of these projects, the sustainability of these projects, and the actions of the Bank's staff and its clients. While attribution is important in nearly every rating, the evaluations also reflect many of the limitations and issues that evaluators identified in the interviews, importantly that the lack of baseline data makes it difficult to prove attribution in practice.

The more surprising findings were related to the logic by which the Bank determines its ratings and how evaluators justified positive and negative ratings in each category. I found that project-level evaluations are, as the evaluators noted in the interviews, a data-driven review, and that many of the ratings are justified by stating outputs or other quantitatively derived facts. Yet, in the absence of this data, many of the ratings are, curiously, justified by stating the operation staff's data as 'fact' rather than considering other forms of data or opinions, including the perceptions of the project's beneficiaries. Another unanticipated observation is that the determining factor that the evaluators consider in assigning a particular rating often has to do with the mandates of the Harmonized Criteria rather than an overarching assessment of the project (World Bank OED, 2005a). In the end, this suggests that the meaning of 'success' in terms of the Bank's projects is more complicated than whether the project had a positive rate of return. This analysis demonstrates the need for further qualitative research to compare and contrast how these ratings are explained and justified in other project categories and over time in order to better understand the full spectrum of how the concepts of ' success' and 'aid effectiveness' are being produced by the Bank in its evaluations.

Chapter Seven

The scale of success in international aid:

Measuring success and defining attribution in the World Bank's country-level evaluations

7.1. Introduction

The literature on the micro-macro paradox (Mosley, 1986; Subramanian and Ragan, 2008; Arndt, Jones, and Tarp, 2010) is based on two assumptions: first, that the impact of the Bank's work in a country would lead to economic changes and, secondly, that there is a clear relationship between the combined rate of return of the Bank's projects and country-level economic growth. In this chapter, these premises are challenged by investigating how the Bank assesses its impact at the country-level and the links between project and country-level impacts. I specifically analyse how the Bank's evaluations measure the impact of its work at the country-level, using a sample of evaluation documents from the region (Africa) in which the Bank has completed the most country-level evaluations. This chapter builds on the analysis of the project-level project validations (called ICR Reviews) in the previous chapter, in which I studied how evaluations determine 'success' in aid projects the logical framework, indicators, and attribution in Bank evaluations at the project-level. In both chapters, I discuss how the format and language of the Bank's evaluations reflect many ways in which the logical framework and data available to evaluators shapes the overall narrative that the Bank 'produces.'

The issue of 'scale' of evaluation, or the geographical scale of the intended impact of the Bank's project or programme, is important to the analysis here. The examination is based on three questions, which involve similar issues as the analysis in the previous chapter on the Bank's project-level evaluations and data. First, how does the Bank's logical framework shape the information about the Bank's impact on countries that evaluations provide? Secondly, what goals does the Bank set for itself on the country-level and what data does it use to prove if these goals have been met? Finally, how do the Bank's evaluations describe and analyse the link between the Bank's ability to meet its goals and the impact of aid on the country-level? (Maxwell, 1997; Cracknell, 2000; Iverson, 2003; Bourguignon and Sundberg, 2007)

The analysis of what 'success' means at the country level and how it is determined in the Bank's evaluations ties all of these questions together. Rather than being focused exclusively on economic growth, I find that the Bank's country-level evaluations often assess the Bank's progress in achieving other aims, especially those related to poverty alleviation. In determining whether these goals have been met, evaluators use both 'top down' indicators and 'bottom up' project results to show the sum of the Bank's work quantitatively. 'Top down' indicators are national-level indicators, such as the poverty level in the country, whereas 'bottom up' indicators are generally the combined results of the Bank's projects on a local level. The current framework does not require evaluators to analyse the link between the Bank's projects and its country-level results, which has important implications for the understanding of attribution in practice (Rebien, 1996). In the end, the investigation in this chapter challenges the conclusions of the micro-macro paradox (Mosley, 1986), research which

investigates neither the Bank's many non-economic impacts nor the complexity of the relationship between the project ratings and the Bank's impact at the country-level.

7.2. Methodology

In order to closely consider the specific language that evaluators use in country-level evaluations, this analysis centres on the data available in the Bank's evaluations in African countries, the region in which the Bank has completed the most country-level evaluations. Using NVivo, a qualitative data analysis programme, I investigated both the framework of these evaluations as well as the specific language used to describe the impact of the Bank's work at the country-level. Given that the country-level evaluations are not uniform in their formatting, it was important to understand the logical framework followed in these kinds of evaluations by creating 'nodes' within NVivo helped to capture the common categories by which evaluators evaluated a country.

After determining the steps by which the Bank determines success and the logic of the assessments, it was critical to examine the logic and language that is included in these evaluations. Looking within these nodes of the common categories within the evaluations, I then dissected how evaluators describe the goals of the country-level evaluation process, how they use indicators, and how they describe the projects and other work that the Bank completed in the country being evaluated. Finally, analysing how evaluators connect their data about the Bank's projects in that country with the outcomes at the country level helped shed light on how evaluators think about the issue of attribution in practice.

7.2.1. Description of the sources used

Before explaining the specific documents upon which the content analysis is based, it is critical to clarify the types of country-level evaluations from which this sample is drawn. Three types of country-level evaluations that the Bank completes on its work are used in this analysis: the Country Assistance Strategy Completion Report (CASCR) Reviews, Country Program Evaluations (CPE), and Country Assistance Evaluations (CAE). These country-level evaluations are a review of the progress of the Bank's work in that country over a defined period of time. According to the Bank, country-level evaluations are results-focused and include a framework of targets and indicators to monitor progress toward the stated outcomes (World Bank, 2011a). To provide an overview of the process, Figure 7.1. summarises the process of country-level evaluation at the Bank that is then detailed in the subsequent paragraphs. The dark boxes at the top of the figure represent the steps that operations team and the International Financial Corporation (IFC) take and the lighter boxes at the bottom of the figure summarise IEG's actions in the evaluation process.

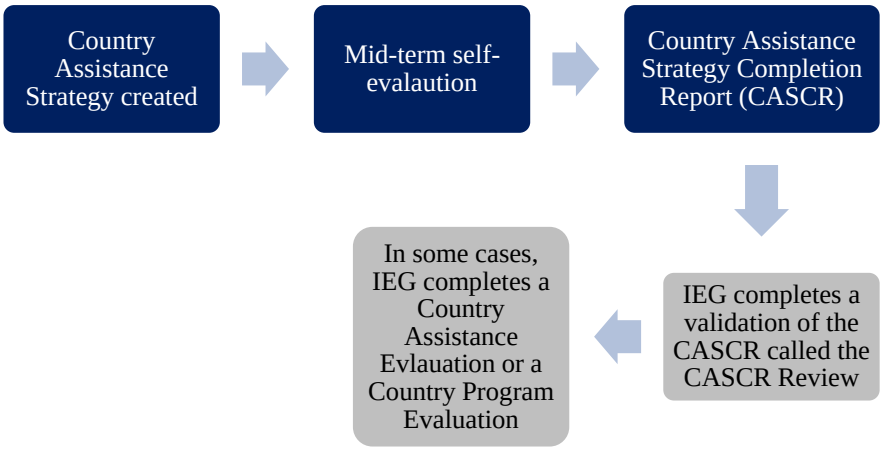


Figure 7.1.: Country-level evaluation process (chart created by author, information drawn from World Bank, 2011a; Interview, Current World Bank Evaluator).

From the Bank's current materials and from the interviews with evaluators at the Bank, it is not apparent why CPEs are completed in some countries and CAEs are completed in others. One evaluator noted in the interviews that it may have to do with client requests for a particular evaluation or a programme from which the Bank can learn important lessons, but this selection process is not clarified or explained in the written evaluations themselves.

The Bank's process at the country-level begins when its operations staff and the International Financial Corporation (IFC), the private-sector arm of the Bank, prepare a Country Assistance Strategy (CAS) to plan a strategy for support linked to the country's development strategy. This document analyses the development challenges facing the country and identifies the key areas where the Bank can have the most efficient impact. Since they began to be implemented in 1990, the country assistance strategy (CAS) has specified the level and type of assistance the World Bank Group will provide to a country, usually for a four-year period. Ideally, the CAS aligns a country's development priorities to the Bank's projects in the country and is meant to be created with the input of the country's leadership (Bretton Woods Project, 2010). The Bank's operations staff members complete a mid-term assessment and then a completion report Country Assistance Strategy Completion Report (CASCR), which is a self-assessment of their work (Bretton Woods Project, 2010).

The Independent Evaluation Group (IEG) enters the project cycle at this phase and completes a review of the self-assessment, Country Assistance Strategy Completion Report Review (the CASCR Review). In some cases, the IEG also completes more in-depth analyses and evaluations at the country-level as well, called the Country Program Evaluations (CPE) and Country Assistance Evaluations (CAE). CAEs examine the World

Bank's performance over the past four to five years in a country and compare these actions with the Bank's promised Country Assistance Strategy (CAS) (World Bank OED, 2005b). Country Program Evaluations (CPEs) are another type of country-level evaluations and are more wide-ranging, considering the scope and impact of the Bank's work over a longer period of time, usually in response to a particular critical event such as the end of a civil war or a financial crisis (World Bank, Appendix B). Given the wide scope of possible development activities in a country, IEG rates only the outcome of the Bank's programmes in CPEs, not the country's overall development outcome (World Bank, Appendix B).

While CASCR reviews are formatted in a fairly standard way, CAS and CPE reports are inconsistent in terms of their formatting. The CASCR Review is similar in format and logic to the project-level validations: the CASCR Review is a verification of the operation team's self assessment and, in the review, the evaluator analyses the Bank's performance in managing the program as described in its strategy document, the Country Assistance Strategy (World Bank, CASCR Guidelines). The CASCR is a text document that has specific sections which must be completed by the evaluators.

The country-level evaluations (the CAE and CPE reports), in contrast to the more limited scope of the CASCR evaluations, examine the performance of a country over a longer period of time but also focus on the impact of the Bank's work in that particular country (World Bank, Evaluation Categories; World Bank, Appendix B). The CAE and CPE reports are more diverse in format and in terms of the visuals they include: CAE reports tend to be non-stylised in format, with the information presented in text and tables, but some CPE reports have the style of a publication, with photos of particular

projects they highlight and stylised formats that are more visually interesting than the CASCR reports, which are published in a form-style with few graphics.

7.2.2. Defining the sample

In the content analysis, I draw on a sample of forty-five of the Bank's country-level evaluations. These country evaluations are all of African countries, the region in which the Bank has completed the most country-level evaluations (please see Appendices 5 and 6 for the list of the evaluations used in this content analysis). The Bank's evaluations that were completed in African countries are the focus in the sample for two important reasons. I chose evaluations from African countries to include in the sample of evaluation documents because Africa is the region in which the Bank has completed the most publically available country-level evaluations. As Table 7.1 below reflects, within African countries, 17 of these country-level reports are CASCR Reviews and 28 are CPE and CAE reports.[2] In this study, the focus is on the way in which the Bank evaluates the development outcomes at the country-level. Given this specific scope, I analyse evaluation process itself rather than the different realities of African development between countries, which varies significantly due to the diverse historical conditions and experiences with structural adjustment programs in African countries (Mkandawire and Soludo, 1999).

[2] I have not used the Country Policy and Institutional Assessment (CPIA), which are not reflective of the outcomes of the Bank's projects, but rather of the country's current policies and institutions, according to the Bank. The Bank uses CPIAs to determine loan sizes (World Bank, "Country Policy and Institutional Assessment, CPIA"; Dollar and Levin, 2005).

Table 7.1: Publically available World Bank country-level evaluations, as of June 2014[3]			
By evaluation type	173 Total Evaluations	121 Country Program Evaluations (CPEs)	52 CAS Completion Report Reviews
Top three regions	49 Africa 34 Latin America and Caribbean 22 East Asia and Pacific	30 Africa (28 usable) 25 Latin America and Caribbean 15 East Asia and Pacific	18 Africa (17 usable) 13 Europe and Central Asia 9 Latin America and Caribbean

I also selected evaluations in African countries for the sample for the content analysis because the original micro-macro paradox research used an African country (Kenya) as an example of aid's different impacts at the project and country-level (Mosley, 1986).[4] Considering the World Bank's data on projects in Kenya in the 1960s and 1970s, Mosley wrote that "[t]here is no significant relationship, positive or negative, between aid inflows measured as a percentage of GNP and the growth of the Kenyan economy over l966-82" but, yet, "projects in Kenya over the period show respectable, if declining, rates of return" (Mosley, 1986, 24-25). The analysis in this chapter focuses on similar issues to the ones that Mosley covered (the project-level and country-level results of international development and the link between the two).

[3] My search on the Bank's website yielded 18 documents in the category of CASCR Reviews and 30 documents in CPEs for Africa. However, not all of these documents are available and a few included in the tallies are repeats of English reports in different languages. As a result, my sample is 28 CAE and CPE reports (the Cote d'Ivoire and Mozambique have duplicate French versions) and 17 CASCR reports (the CASCR for Burkina Faso was not publically available at the time of writing).

[4] The country evaluations that are publically available consider the Bank's work in 107 countries: only 16 countries, including Kenya, have three or more country-level evaluations. Brazil has eight country-level evaluations, the most of any country.

7.3. The Bank's logical framework at the country-level

The logical framework ('log frame') of evaluations in international development organisations such as the World Bank determines what kind of information evaluators use to base their decisions. The analytical approach in this chapter was influenced by the previous research on the theory of monitoring and evaluation (Maxwell, 1997; Weiss, 1997; Grove and Zwi, 2008). I analysed both what the Bank defines as its impact on the country-level and the methods by which the Bank determines its impact, especially which factors the country-level evaluation framework prioritises within the institutional context of the Bank. The findings are consistent with those of previous studies that argue that the Bank's evaluations follow a technocratic logical framework (Pawson, 2003; Weiss, 1997), which has particular implications for how the Bank considers the impact of its work on the country-level. The Bank's log frame shapes its evaluations by prioritising quantitative indicators and excluding other considerations of aid impact and beneficiary feedback. Overall, the log frame produces an understanding of the Bank's impact at the country level that is distilled into the sum of its quantifiable project outputs. The evaluation process by which the Bank determines the impact of its work within a country is important to understanding how 'success' in aid is defined for recipient countries.

The steps that the Bank takes in defining its country-level impact need to be explained in order to explore the process. A preliminary problem in analysing country-level evaluations is that they are not as uniformly formatted as project-level evaluations are. On face value, the country-level evaluations assume different formats, but, on closer inspection, they do generally follow common steps. I coded the common steps that the evaluators take in these evaluations and found that the Bank's country-level evaluations present and evaluate the Bank's work in the country in similar ways across the data in the

sample, analysing the goal, purpose, output, activities of the work and the indicators. At the beginning of the evaluation, the evaluator reviews the country's history and current development state, identifying the key development issues and the historical and other realities that mitigated the success of the Bank's work, such as civil war or other external challenges. Next, they describe the World Bank's strategy in the country and the 'pillars' (or goals) that the Bank staff pursued, which are often those identified in country-led initiatives such as poverty reduction strategy papers (PRSP).

After identifying the pillars or goals of the Bank's work in the country, the evaluator analyses the extent to which the goals have been met, drawing on both 'top down' indicators such as the country's position in relation to the Millennium Development Goals (MDGs) and 'bottom up' indicators such as the ratings of project-level results. The specific way in which indicators are used and how these types of indicators are connected will be described later in this chapter. In the CASCR Reviews, as at the project-level evaluations, IEG often evaluates the outcomes and other variables along six rating categories from highly satisfactory to highly unsatisfactory (World Bank IEG, Appendix B), which are derived from international standards (World Bank, Methodology). In some cases, the CAE and CPE reports provide ratings for the Bank's work in each particular area, but this is not always the case. At the end of their evaluations, the evaluators identify the strengths and weaknesses of the Bank's and the recipient's roles in the development and also identify lessons learned.

7.3.1 The Bank's logical framework in context

The log frame shapes the information evaluators provide about countries in several critical ways, most importantly, that impacts must usually be quantifiable in order

to be included in this log frame and that the methodology does not require the feedback of beneficiaries in the overall assessment. The Bank's log frame in many ways reflects the form and limitations that previous studies about monitoring and evaluation have identified. Previous studies in the field of monitoring and evaluation have argued that the log frame that the Bank and other large aid donors use obscures less quantifiable relationships and the positive and negative consequences of aid projects (Grove and Zwi, 2008; Maxwell, 1997; Rebien, 1996). As in the project-level evaluations, the Bank's country-level evaluations use an evaluative framework that relies heavily on quantitative data. Evaluators' written comments in the country-level evaluations reflect that the critical task of the evaluator in the context of the Bank is to assess the data provided rather than to explain the impact of aid comprehensively. As one evaluator summarised the way in which they reached their overall determination:

> The CAS results framework is clear, but the outcome indicators are in some cases very partial proxies of the outcomes sought. For example, the business environment indicators (on government arrears and delays in commercial court case adjudications) are inadequate measures of the overall business climate (Burundi, CASCR Review).

The focus on indicators and measurements in this quote reflects how evaluators assess the Bank's country-level impact in many evaluations in the sample. Their determinations are based on the extent to which certain targets are met rather than the unintended or other consequences of the Bank's work in that country. An important implication of focusing on targets rather than the specific implications of each decision in the programme is that it does not require the evaluator to consider the elements of a programme that led to specific results and why these results are observable. In all of the evaluations that were analysed, there was no evaluation that explicitly traced the choices

made about a particular project or country programme to the outputs and the indicators it claimed to reach.

Given the similar conclusions in previous studies that consider log frames in development evaluation, it is unsurprising that the Bank's country-level evaluation system values almost exclusively quantitative data. In focusing on project results rather than specific examples of beneficiary or in-country feedback, it encourages a technical approach to evaluation, which previous studies argue may be influenced by the ideas of New Public Management, a private-sector approach (Blalock, 1999). As Holvoet and Renard articulated:

> One of the consequences of this leaning towards monitoring at the expense of evaluation is that the focus is much more on pure stocktaking of performance *('were the targets met?')* than on probing into underlying reasons for (non)-performance *('why were the targets not met?')* (Holvoet and Renard, 2007, 71).

Based on the literature regarding monitoring and evaluation, there are certainly alternative methods by which the Bank could consider its impact at the country-level that would broaden and deepen its conclusions. Despite the rise in prominence of impact evaluations in international development (Banerjee and Duflo, 2011), only a few of the country-level evaluations in the sample referenced in this chapter mention the concept of the 'counter factual' as a part of the way in which the data about the country's progress was considered. For example, as one evaluator clarified, a counterfactual approach to country-level evaluations would ask the questions such as:

> What could the Bank have done differently? If the Bank had analysed the extent of country ownership of reforms, based on the degree of consensus among decision makers, the locus of policy initiatives (whether local or external), the existence of upfront actions, and the participation of major stakeholders, it would have realized that ownership was weak, and it would not have relied so heavily on commitments with technocrats in the ministries (Zimbabwe, CAE, 2004, 26).

These questions about the choices the Bank and its staff made during their work in a country are critical to understanding the nature of the Bank's impact on a country, but are often not considered within the context of country-level evaluations. In the above quote, the evaluator asks what the Bank could have done differently, but the analysis focuses instead on the quantitative results that are actually reported.

An additional implication of the Bank's evaluative framework at the country level is that it does not require evaluators to consider the feedback of beneficiaries in an engaged way. Given the previous research on this topic (Chouinard, 2013), this lack of attention to the beneficiaries' feedback was generally expected. In fact, none of the evaluations I reviewed quotes directly from the views of the beneficiaries of the Bank's projects. One evaluation does refer to a poverty assessment which aimed to find a "consensus on the definition of poverty and actions to alleviate it" using participatory poverty assessments and surveys to gauge the cost of living in various areas of the country. This is, however, the only participant-driven process referred to in the country-level evaluations I reviewed. The Ethiopia CAE mentions the client's perception ("only 15 percent responding favourably to the question about the Bank's impact on the status of the poor") but the client, in this case, appears to be the high-level government officials rather than the beneficiaries of the project aimed at low-income communities (Ethiopia CAE, 2000, 7). Three of the evaluations refer to household surveys, but two of these also note that the results of household level surveys are not available (Mozambique and Cameroon).

By not considering the viewpoints of beneficiaries, the Bank does not provide a context about the meaning of the results that are meant to improve the lives of the people within the client country. A review of the literature on monitoring and evaluation reflects

that there are other ways in which the Bank's log frame could engage with the impact of its projects or its overall impact within countries. A 'theory based evaluation' approach, in contrast, would seek to find the cause and the effect of aid and test the assumptions about causal relationships in aid programs rather than just stating data. In addition, theory based evaluations and other models that use more qualitative data include and require space for the viewpoints of the beneficiaries to be considered in the overall evaluation of an aid programme (Weiss, 1997; Pawson, 2003).

7.4. The Bank's goals at the country-level: What and how?

Country-level evaluations measure the Bank's impact in the country based on the goals that the Bank's operations staff, who implement the projects, sets for themselves, not the overall economic progression of that particular country. It is, therefore, important to understand both how evaluators judge the Bank's impact overall and how they assess particular goals. The analysis below is divided into analysing 'what' goals these evaluations focus on and 'how' the evaluations assess the Bank's impact in these areas within countries.

7.4.1. What?: The Bank's goals at the country-level

The Bank's stated goals are now much more focused on the Bank's impact on poverty alleviation than when Mosley wrote his original study on the micro-macro paradox (Mosley, 1986), reflecting trends in international development aid (World Bank 2009; Stiglitz et al., 2009). The Bank determines whether or not it has completed these goals by using indicators to illustrate the overall situation within the country or by

providing a sum of the Bank's project results in order to summarise its impact. The Bank's evaluators describe the extent to which the Bank met its own goals through quantitative means rather than discussing the specifics of the Bank's work that led to particular outcomes.

I begin by analysing what the Bank's goals are at the country-level and how its evaluators assess these goals. The contemporary studies on the micro-macro paradox (see, for example, Arndt, Jones, and Tarp, 2010; Howes, 2011) continue to compare the percentage of successful projects within a country to the overall economic changes within the country. The first critical point is that many of the indicators in the Bank's country-level evaluations are not necessarily related to economic growth but rather to other goals such as poverty alleviation, human development, and governance. This suggests that the goals against which the Bank considers its overall impact on a country have changed significantly since Mosley wrote his article on the micro-macro paradox in 1986, when the emphasis of the Bank's work was primarily on economic development. This reflects some of the recommendations for the goals of aid articulated in the Stiglitz report (Stiglitz et al., 2009). As the below table reflects, although economic growth is the most common theme along which most countries are evaluated, the single most common pillar as stated in the Bank's evaluations is categorised as "human development and poverty reduction" (22 of the 45 evaluations included in the sample, please see Appendix 7 for further detail).

Table 7.2.: Pillars in the evaluations included in the sample, by theme and title				
Theme	**Pillar**	**CPE/CAE**	**CASCR**	**Total**
Economic growth	Pro poor growth	2	2	4
	Macroeconomic stabilisation	13	2	15
	Financial reforms	2	1	3
	Sustainable growth	2	4	6
	Broad based growth	3	3	6
	Private sector economic growth	9	6	15
	Labour intensive growth	1	0	1
	Economic growth, Total	**32**	**18**	**50**
Poverty and livelihoods	Human development/ Poverty reduction	18	4	22
	Deliver social services, community empowerment	8	13	21
	Refugees	1	0	1
	Poverty and livelihoods, Total	**27**	**17**	**44**
Government	Governance	5	9	14
	Public finance management	3	0	3
	Institutional development	5	4	9
	Public sector restructuring	4	0	4
	Budgeting	2	1	3
	Government, Total	**19**	**14**	**33**
Infrastructure	Infrastructure and development	4	1	5
	Infrastructure, Total	**4**	**1**	**5**
Sector reform	Agriculture	4	1	5
	General reform	2	1	3
	Transportation	1	0	1
	Sector reform, Total	**7**	**2**	**9**
Monitoring and Evaluation	General Monitoring and Evaluation	0	1	1
	Monitoring and Evaluation, Total	**0**	**1**	**1**
Most common pillar	**2nd most common pillar**			

The distribution of these goals reflects the context in which the Bank's staff articulated them: Of the forty-five evaluations included in this sample, forty-one were completed in 2000 or later and review the Bank's work completed in the mid-1990s and beyond. These evaluations, therefore, reflect the important changes within the Bank and in the wider development donor community about the goals of aid since the 1990s (Sen, 1999;

Stiglitz, 2009). Evaluation is an ever-changing field and, while the framework is still technocratic, there have been significant changes since the 1990s in how aid donors that are similar to the Bank think about their work. In many cases, the evaluations note the fact the Bank's goals are changing, for example:

> Early in 1994, the emphasis of Bank assistance strategy shifted toward assisting Zambia in achieving sustainable economic growth and reducing poverty. ... While accepting the need to continue support for structural adjustment and to help close the massive financing gap, greater priority was given to improving the enabling environment for private sector growth through investments in infrastructure and human resources and to targeting the poor and vulnerable groups with specific programs in agriculture and social services (Zambia CAE, 1996, section 6).

The changing goals that the Bank sets for itself in its client countries reflect changes in the goals of aid in the wider context of international development. These changes have been signalled by the Bank's acceptance of certain international agreements, including the Poverty Reduction Strategy Paper (PRSP), the acceptance of the Millennium Development Goals (MDGs), and other aid declarations such as the Paris Declaration (Conlin and Stirrat, 2008). These international agreements suggest how the economic production and distribution in development has changed over time and focus on the potential of using aid to alleviate poverty and the importance of country ownership. The Poverty Reduction Strategy Paper (PRSP) approach, in particular, has shaped how goals are determined and assessed. Rather than the Bank deciding upon the goals for its projects within the client country, since the 1990s, the goals are articulated in the PSRP strategy documents which are drafted by in-country leadership in collaboration with the Bank and, thus, reflect some in-country stakeholder's views of what the aims should be (Holvoet and Renard, 2007). The adoption of the PRSP shaped the evaluation teams in important ways because it required increased data collection and encouraged

client countries to develop monitoring and evaluation systems which hold governments accountable by providing reliable data (Wood, 2011).

7.4.2. How?: Assessing the Bank's impact on poverty alleviation

In addition to clarifying what the Bank's goals are at the country-level, it was important to understand the logic that evaluators use to assess the Bank's success in reaching its goals. In this section, I analyse how the Bank's evaluations explain its impact in terms of the goal of 'human development and poverty alleviation,' listed as a 'pillar' in 22 of the 45 evaluations of African countries included in this analysis. In the evaluations, it is clear that the goal of the evaluation is not just whether poverty increased or decreased but, rather, what, in particular, the Bank's impact was on issues related to poverty within the country, however limited that scope may be.

In assessing the Bank's impact in terms of poverty alleviation, the Bank uses data from its projects that it views as being related to poverty alleviation rather than only focusing on whether poverty overall had decreased. Evaluators often gauge the quality, for example, of the Bank's analytic work, its monitoring and evaluation program, and its efforts to encourage policies that direct efforts towards the poor. To demonstrate this, evaluators use indicators from both the Bank's project data and international poverty reduction indicators. In the written evaluations, evaluators do not, however, investigate reasons for the Bank's impact or the meaning of these quantifiable indicators for the country and its people. Rather, evaluators rely on quantitative indicators to show the sum of the products of the Bank's work rather than more comprehensive assessments.

In the evaluations, evaluators often measure the Bank's contribution to 'poverty reduction and human development' as a goal by assessing the combined results of its

projects within the country. As is explained above in section 7.3.1., evaluators use both 'top down' indicators such as the country's position in relation to the Millennium Development Goals (MDGs) and 'bottom up' indicators such as the ratings of its project-level results. The project-level results are formed from the Bank's evaluation processes which were discussed in the previous chapter. The project-level data which forms the 'bottom up' indicators is important to consider as this is the data upon which the Bank's overall determinations of its impact are based.

Table 7.3. below summarises the distribution of the kinds of projects that are referenced in the Bank's evaluations to assess the Bank's contribution to alleviating poverty in its client countries. As the table demonstrates, health and education are the two most common project areas in the evaluations in this sample that the Bank references in the evaluations in determining the extent of poverty alleviation. How evaluators present the data from these projects is important to understanding how 'success' is determined at the country level. In describing the results of the Bank's education and health projects, evaluators focus on quantitative indicators, using project level results in sum to show the Bank's progress toward a particular goal. This is similar to the strategy and approach the evaluators use in project-level evaluations, as I discussed in the previous chapter.

Table 7.3.: Projects that evaluators use to assess poverty reduction goals	
Areas related to human development and poverty reduction goals	**Countries[5]**
Health	Uganda (CAE, 2009) Mali (CAE, 2007) Madagascar (CAE, 2006) Malawi (CAE, 2006) Mauritania (CAE, 2005) Lesotho (Joint Evaluation with ADB, 2002) Ethiopia (CAE, 2000)
Education	Uganda (CAE, 2009) Mali (CAE, 2007) Madagascar (CAE, 2006) Mauritania (CAE, 2005) Lesotho (Joint Evaluation with ADB, 2002) Ethiopia (CAE, 2000) Burkina Faso (CAE, 2000)
Poverty reduction and human development (general)	Mozambique (CPE, 2011) Kenya (CAE, 2004) Cameroon, (CAE, 2001) Malawi (CAE, 2000) Ghana (CAE, 2000) Ghana (CAE, 1996) Eritrea (CAE, 2004)
Infrastructure	Malawi (CAE, 2006) Mauritania (CAE, 2005) Ethiopia (CAE, 2000)
Water and sanitation	Uganda (CAE, 2009) Madagascar (CAE, 2006) Republic of Congo (CASCR Review, 2012)
Protect vulnerable groups	Madagascar (CAE, 2006) Ethiopia (CASCR Review, 2012) Malawi (CAE, 2006)
Women and girls	Senegal (CAE, 2006) Mauritania (CAE, 2005) Burkina Faso (CAE, 2000)
Equity and poverty, shared growth	Zimbabwe (CAE, 2004) Senegal (CASCR, 2013)
Raise living standards	Mali (CAE, 2007) Senegal (CAE, 2006)

[5] Other issues which only appeared in one evaluation included: Urban and rural development (South Africa CASCR Review, 2013); Agricultural growth (Mozambique CAE, 1998); HIV/AIDS (Malawi CAE, 2006); Slow population growth (Malawi 2000); Food security (Malawi CAE, 2006); Increase pro-poor expenditures (Madagascar CAE, 2006); Decentralise institutions (Senegal CAE).

As is anticipated by the theory of monitoring and evaluation (Cracknell, 2000), evaluators 'prove' that poverty reduction pillars have been reached using similar logic. In education, evaluators focus on the data which show that the Bank has expanded access to education, funding has increased as a result of the Bank's policies, and literacy rates have also improved. Many evaluators focus on the Bank's impact on girls' education in particular, which the Bank often emphasises is an important part of a strategy to alleviate poverty. As one evaluation notes in a written assessment:

> The gross enrolment ratio reached 92 percent in 2002/03, a significant gain, though below the target of 100 percent by 1999/2000. The focus on girls' education permitted the rise in the enrolment of girls to 49 percent of total enrolment in 2001 from 46 percent in 1996 (Mauritania CAE, 2005, 18).

In explaining the Bank's successful impact in education, they often note not only the Bank's ability to reach its target, but also how quickly the Bank progressed in reaching this goal. In both ways, this allows the evaluator to quantify the Bank's progress towards the goal but does not require a qualitative assessment of the impacts of what this progress may be.

Evaluators also use quantifiable indicators to determine the impact of health projects, such as the number of health facilities the Bank constructed, the country's HIV/AIDS rate, infant and mother mortality rates, and the increase of basic health services as a percentage of population. Evaluators do not, however, consider the implications of these project results and, as in education, rather only present these quantifiable benefits as proof that the Bank has made progress and a positive impact in alleviating poverty by improving key health indicators. As another evaluation notes:

> Together with other development partners, the Bank successfully supported the construction of numerous health facilities, improved a number of institutions, trained many health professionals, and contributed to increased health service access. This in turn led to improvements in key health indicators, including infant and maternal mortality rates (Mozambique CPE, 2011, xiv).

Interestingly, here and in a few other evaluations in the sample, the Bank's evaluator draws a direct link between its project outputs and related indicators within the country but does not explain how these particular projects led to these particular outcomes within the country, in this case Mozambique. This quote highlights additions to the health care system in Mozambique at the country level, but does not question the quality of these benefits. The number of institutions and health care professionals, for example, are given as evidence of increased health care in Mozambique, but the evaluator provides no analysis of the quality of these contributions or how they specifically contribute to improvements in Mozambican health care.

Overall, as the previous examples cited in this section reflect, in considering the Bank's progress towards its self-defined goals, evaluators generally provide the sum of the quantitative indicators of the Bank's projects in that particular area: the total girls in schools reported in each of the Bank's education project, for example, is provided as the key evidence of the Bank's impact on education. The evaluators, however, are not required to explain what the Bank's specific contribution is or consider the quality of these contributions beyond the quantitative measurements.

7.4.3. From goals to country-level impact

The aim of this section is to analyse the logic by which the Bank's evaluators use the data to determine the extent to which the Bank achieved particular goals and thus to demonstrate the Bank's impact on the recipient country. Given the research question

about ‘scale’ in aid in the Bank’s evaluations, it was critical to examine how evaluators connect the results of projects or groups of projects (the ‘bottom up’ indicators) to the Bank’s impact at the country level (the ‘top down’ indicators). In general, in their discussion within the written evaluations, evaluators often connect the results of the goals within the countries (for example, the number of hospitals built within education projects) to the overall determination of the Bank’s success within a country. As described above, however, rather than investigate how particular projects led to particular outcomes within the client countries, the Bank’s evaluators often only combine the results of these projects within areas such as health and education and projects to determine the Bank’s overall impact for the relevant pillar within the country.

I coded the data within the Bank’s evaluations included in the sample that list ‘poverty and human development’ as a goal to determine how evaluators assess the Bank’s overall impact and the language they use to describe this impact. There are several common ways in which evaluators present indicators to explain the Bank’s impact on poverty levels: they state the change in the poverty level within the country, they state the rate of poverty as a part of a table that shows how one country compares to other sub-Saharan African countries, and, in the absence of a figure on the overall poverty levels, they provide other indicators such as, for example, increases in sanitation access.

How the data are used is important to consider (Merry, 2011), because the context suggests the logic by which these judgments are determined. In the evaluations, the evaluator states the data or the ability of the Bank to meet its indicator as definitive proof of success, often without further analysis. As the next section will discuss, the link between the Bank’s project results and it impact within the country is ill-defined in the

evaluations, in part because the evaluations do not present a comprehensive evaluation of the specific characteristics of projects or groups of projects that allowed for and enabled their positive ratings.

As tables 7.4. and 7.5. below reflect, in determining the Bank's impact, the evaluators also consider other indicators outside of the data provided in the project-level results. In their evaluations, Bank evaluators note that the Bank does this review through both a 'top-down' review, which focuses on whether the Bank program achieved its outcome and had an impact on the country's development, and a 'bottom-up' review of whether the Bank's services achieved the objectives. The Bank claims that these approaches test the consistency of findings from the products and services and the development impact (World Bank IEG, Appendix B). As the next section will discuss in further detail, however, there is little discussion of the link between the two levels of indicators.

The three examples in the first table below summarise how the Bank uses top down and international indicators to demonstrate the impact they have had in a country. On the 'top down' level, evaluators often list the national-level statistics such as the poverty level without clarifying how the Bank's work specifically contributed to that result. For example, evaluators use Transparency International indicators that measure corruption in their assessment of governance pillars and Millennium Development Goals for pillars related to poverty. If they use top down indicators, evaluators rarely discuss the details of the Bank's projects, often presenting project data in a chart at the end of the report without discussing the contribution of these specific projects.

Table 7.4.: Top down indicators used in the Bank's evaluations	
How the Bank describes its overall impact	**Example**
State the level of poverty and the changes over time	"The incidence of poverty declined impressively from 69 percent in 1997 to 54 percent in 2003, whereas poverty remained higher in rural areas (55 percent) than in urban areas (52 percent)" (Mozambique CPE, 2011, xiv).
Show progress in Millennium Development Goals (MDG) and other indicators that are related to poverty	"Bank support focused on helping Senegal achieve universal primary education and the related MDG of gender parity increased from 58 percent in 1990/91 to 80 percent in 2002/03, while the average net enrolment increased from 47 percent to 58 percent and the ratio of girls to boys almost doubled. Secondary school enrolment also rose, though at a more moderate rate" (Senegal CAE, 2006, 27).
Compare the country's indicators to other sub-Saharan African countries	"In infant mortality, Uganda ranked squarely in the middle after Ghana, Kenya, and Tanzania. However, the country made the greatest gains in immunization rates for diphtheria and measles during 1986–95. Uganda has also progressed in education. It raised its rank in gross primary enrolments from sixth among the seven countries in 1980 to fourth by 1995, even before universal primary education was introduced. Uganda also made great strides in overcoming illiteracy. But it ranked lowest on life expectancy and the death rate" (Uganda CAE, 2001, 25).

Some evaluations use the Bank's projects and the sum of their impact within the country to prove the Bank's impact; these are referred to as 'bottom up' indicators. As in the evaluations that use 'top down' indicators, evaluations that use bottom up data have a dearth of discussion about how these projects led to specific outcomes in the country. As the two examples in the below-table reflect, instead the evaluator often reports the areas that the Bank worked in and the positive outcomes of the Bank's projects in these areas. Another way in which evaluators provide evidence for the Bank's impact was to report the percentage of successful projects in the country supported by the Bank.

Table 7.5.: Bottom up indicators used in the Bank's evaluations	
How the Bank describes its overall impact	**Example**
Refers to the positive outcomes of the Bank's projects in particular areas	"The Bank is supporting direct actions to reduce poverty through the programs on education, health, and roads, and the social fund project. … The general poverty orientation of the Bank's current portfolio is satisfactory" (Ethiopia CAE, 2000, 7).
Refers to the percentage of successful projects in the countries' portfolio	"Out of the 23 ongoing projects at end February 2000, only 5 were rated unsatisfactory, representing 8 percent of commitments and 11 percent of the undisbursed balance" (Ghana CAE, 2000, 11). "The outcome of projects approved during this period was rated satisfactory for 87 percent of the commitments, and the country's development objectives were broadly in line with the Bank's corporate objectives. The outcome of the program through 1981 is therefore rated satisfactory" (Cameroon CAE, 2001, Memorandum).

It is not clear why some evaluators focus more on 'top down' indicators and others utilise 'bottom up' indicators in the country-level evaluations. Based upon the feedback from the evaluators themselves in the interviews, the differences may be based upon differences in the data available to evaluations staff at the time of the data review rather a methodological reason. The important point to note is that, whether they use primarily top down or 'bottom up' indicators, the evaluators specifically use quantitative data to illustrate the current status of certain issues in country, either that it is meeting its internationally defined targets or that the projects within the country have been successful. In general, there is minimal interrogation about what led to the specific results that the evaluators report, whether they are quoting from the Bank's project-level results or comparing the indicators within the region to the indicators found within the country.

Compared to how evaluators assess the impact of the Bank's work on a particular pillar, in assessing the Bank's overarching impact within a country, the written evaluations at the country-level suggest that there are more obstacles in determining the overall impact of the Bank's work due to the scope of issues that must be considered within a country. The majority of evaluations, while projecting an overall positive image of the Bank's impact in the country, report that the programmes had mixed results and impacts at the country-level. For example, in the Mozambique CPE from 2011, the evaluator finds the overall efforts to achieve the pillar of "poverty reduction and human development" as moderately satisfactory, noting that poverty declined from 69 percent in 1997 to 54 percent in 2003 but the growth was unequal (Mozambique CPE, 2011, xiv). The evaluator's comments reflect that there are some obstacles to determining the total impact of the Bank's work in a particular country, particularly because it is difficult to see the impact of these policies within the timeframe of evaluation, a reality which is anticipated by the existing theory of monitoring and evaluation (Bourguignon and Sundberg, 2007).

Overall, the country-level evaluations help to clarify the discernible impacts of the Bank's work, but do not explain how or why certain aims were achieved due to the cumulative results of the particular projects selected. From the analysis of the bottom up indicators, the logic is that, if projects within this category were successful, then that suggests the Bank made a positive contribution towards this pillar. From the analysis of the top down indicators, the logic is that, if the country made progress towards meeting the target or indicator, then the Bank had a positive impact on the country.

7.5. Attribution in the Bank's country-level evaluations

The previous chapter on project-level evaluations included the finding that the evaluations often mention the difficulty in proving attribution, but that the issue of attribution (and the related issues of the 'results framework' or 'causal chain') is often mentioned. Here I review how the Bank's evaluators articulate the link between the Bank's project results as well as the hindrances they often identify in attempting to articulate this connection. The Bank's country-level evaluations do not comprehensively consider the links between project outcomes and the overall impact the Bank had, in part because of the limitations in the evaluative framework at the Bank and in part because, as theorists have argued, it is difficult to prove this link (Iverson, 2003; Cracknell, 2000). In assessing the difficulty that evaluators and theorists perceive in determining the link between project-level and country-level results, I challenge the conclusion of the micro-macro paradox, which is based on the premise that there is a clear link between these levels of aid.

In order to understand how evaluations currently conceive of and explain attribution and the link between the Bank's project level and country-level impacts, I analysed how evaluators linked project results to country-level results, either directly or indirectly. If the evaluators do not analyse the link between the top down indicators and project outcomes, as is the case in the evaluations in the sample used for this analysis, to what are the country-level results they cite attributed? Some evaluations are clearer in terms of attribution, suggesting that there are links between the outcomes perceived at the country-level and the project-level results the evaluator reports. The Bank is clear that it uses both top down and bottom up evaluation indicators, drawing from both its own

project ratings and history as well as comparing the country's economic, political, and social indicators with other countries. As one evaluator wrote:

> The preceding chapter provides a 'bottom-up' evaluation of outcomes for Bank products and services. This chapter assesses development effectiveness by looking at outcomes from the 'top-down' perspective of aggregate performance indicators (Zambia CAE, 2002, 1).

As this quote reflects, however, while there is ample discussion of top down and bottom up indicators separately, evaluators do not explicitly consider the link between the two and what the Bank's specific contributions may be at the country-level. This may be, in great part, because of the difficulty of asserting that the Bank's work is responsible for changes in national-level indicators, especially if it has a smaller portfolio in that country.

In the written evaluations in the sample, evaluators often refer to the Bank's projects and suggest that they should be able to link these results to the country-level. The link between project-level and country-level results, however, is not well articulated. Evaluators may explain, for example, that the poverty alleviation strategy has been implemented through particular project and may then list each of the target projects and what they achieved. In many ways, this vague logical strategy in terms of attribution is not unexpected: In their written comments, several evaluators note that 'attribution is difficult to establish' but then follow this statement by citing indicators regarding issues such as infant mortality or school enrolment. The evaluations, however, do not clearly describe why there might be a link between the poverty alleviation, as cited in the above quote, and the work that the Bank has undertaken in the country. Evaluators' comments reflect that determining whether the Bank's overall progress was positive or not is a difficult process and that they generally attribute country-level results to the Bank's work

only insofar as they are able to provide the sum of quantitative results from the Bank's projects.

As I have discussed, the log frame at the Bank does not require evaluators to consider how the Bank's projects led to particular impacts at the country-level, although many evaluators do allude to how certain project results contributed to outputs at a larger scale. A review of how evaluators describe the process of attribution reveals several impediments to the process, some of which are related to the limitations of the evaluative framework and others of which are related to the difficult nature of promoting development in the real world rather than in theory. In its explanatory materials, IEG notes that the program will be similar to the overall development progress only if a Bank Group assistance program is a significant part of the resources devoted to a particular country. IEG notes that the Bank's "program outcomes sometimes diverge significantly from the client's overall development progress," (IEG, Appendix B, 89) emphasising that a positive program outcome can occur in a country with an overall unsatisfactory development process and vice versa. Considering the apparent contradiction, IEG notes the apparent incongruity "becomes clearer once we consider that the Bank Group's contribution to the outcome of its assistance program is only part of the story" (World Bank IEG, Appendix B, 89) because the complete story includes the impact of the client, other partners, and exogenous forces such as international economic shocks.

This insight into the issues with attribution in the real world are important to consider in the context of the findings of the 'aid effectiveness' literature. The micro-macro paradox literature does not consider the logical problems in defining the link between the Bank's projects and its impact on countries, especially within the Bank's technocratic log frame that only considers quantitative data (Mosley, 1986). The

theoretical research suggests that there is not yet consensus about how the impacts of aid can be attributed to a particular project, the timing of evaluation, and the qualitative and quantitative methods that can be used (Rebien, 1996; Cracknell, 2000).

There are many reasons why evaluators may be unable to make a clear link between project level results and country-level impacts. To begin with, although the micro-macro paradox suggests that there should be a clear link between project success and impacts at the country-level, the theory of monitoring and evaluation has found this link to be much more complicated. The question of attribution and the link between projects and the observable results on a national level is a difficult challenge to evaluators and programmers within institutions that are similar to the Bank (Davidson, 2000) and few studies link variables and programs to determine the reasons for particular results (Weiss, 1997). This is perhaps, as Cracknell suggests (2000, 114), because there is a "degree of unreality about the crude assumption 'if a then b'." Previous studies on the theory of attribution instead generally focus on the 'black box' between a project design and its impacts (Iverson, 2003; Bourguignon and Sundberg, 2007; Cracknell, 2000; Rebien, 1996) and the institutional, political, and other reasons for average project success in some countries rather than others (Dollar and Levin, 2006; Dollar and Svensson, 2000).

Next, in the interviews at the Bank, evaluators suggested that it is difficult to say that successful projects should "add up" because the Bank's role in different countries is varied. For example, as one evaluator articulated, if the Bank's investment in a country is a significant amount of the total foreign investment in the country, then it perhaps makes sense to consider if the projects have had an impact on growth, but it would not make sense if the Bank only contributed a small percentage of the overall aid or funding for

social programmes in the country (Interview, World Bank Evaluator). In the discussion of particular impacts in the written evaluations, evaluators similarly write that there are too many factors at the country-level that are not accounted for in the final rating at the project level and, importantly, that the Bank's role in a country may or may not have an economic impact even it is was important in achieving other ends (Bourguignon and Sundberg, 2007). As one evaluator articulated in a country-level evaluation on the issue of service delivery: "Are these projects helping the community become a more effective unit in driving demand for services from the bottom up? The picture thus far is mixed" (Nigeria CAE, 2010, 78). In addition, often a major problem that evaluators face in determining attribution is the lack of monitoring data, noting, for example, that there was no monitoring system at the beginning of the Bank's work in the country and that they can only estimate the changes in the number of hospitals or pupils due to the Bank's work. As one evaluator pointedly wrote:

> The project was also designed to expand access to and improve the quality of secondary education, improve the quality and management of higher education, and promote private sector provision of education. These objectives were met. However, student learning outcomes, a key measure of the quality of education, were not part of the project design and so it is not possible to tell whether student learning has improved (Mauritania CAE, 2005, 18).

Thus, even if the outcomes have been met, as in this case, evaluators' comments about the logic of proving that the Bank's efforts led to these outcomes reflects a more complicated relationship between the Bank's work and the observable impacts in its client countries.

In the end, the language and logic of the Bank's evaluations, the feedback of Bank evaluators, and the literature on the theory of monitoring and evaluation suggests how the question of attribution in practice is far more complicated than the relationship

expected by the micro-macro paradox literature. As I have argued, the way in which the Bank's evaluations currently use only quantitative indicators to determine whether it has met its goals and the overall impact in the country makes it difficult for evaluators to complete the kind of detailed evaluation that determining attribution would require. In addition, the theorists and the evaluators themselves have highlighted the obstacles to attribution that poor monitoring and theoretical issues pose to making clearer links between the Bank's projects and overall impact in a country possible.

7.6. Summary

How the Bank evaluates its impact on the country-level and the data upon which the Bank's evaluators base these assessments have several critical impacts on the final results of these evaluations. The country-level evaluations help to explain if certain aims were achieved, but do not explain how or why certain aims were achieved with the particular projects selected. This is similar to the nature of the Bank's evaluations at the project-level, where evaluators often consider the issues related to attribution but logically define certain ratings using the operation staff's viewpoints rather than an overall assessment of the project's impact.

These findings, in many ways, reflect the results of the analysis in the previous chapter. In considering project-level data in the previous chapter, I confirmed the evaluators' feedback that the ICR Review is an accountability tool and that it is a 'validation' of the project's stated aims rather than an in-depth evaluation. The Bank's log frame at the country-level, as at the project-level, most often requires evaluators to express the impacts of the Bank's work in quantifiable ways that do not consider the feedback of beneficiaries in the overall assessment. As a result of this log frame, the

Bank's country-level evaluations do not consider the links between project outcomes and the overall impact the Bank had and instead use indicators and project results to show the sum of the Bank's work quantitatively. I also found that the determining factor that the evaluators consider in assigning a particular rating has to do more with the mandates of the Harmonized Criteria rather than an overarching assessment of the project.

The analysis in this chapter both clarifies the nature of the Bank's assessment at the country-level and serves to challenge the previous research on the micro-macro paradox. I found that many of the indicators in the Bank's country-level evaluations are not related to economic growth and that evaluators themselves recognise the difficulty of connecting different levels of aid in the context of the real world, in which there are many variables that these country-level evaluations are not designed to consider. Similarly, at the project-level, there are different factors by which evaluators assess projects, many of which are not related to economic growth but rather to the Bank's country programme delivering what it promised to deliver. In this way, at both scales of aid, the analyses of the processes and institutional role of evaluations as well as the logical constructs of these documents suggests that evaluation at the Bank is more of an accountability mechanism than an assessment of the contribution to economic growth. These findings also suggest the importance of using the full contents of the Bank's evaluations in arriving at a more nuanced understanding of the way in which aid can be effective in the real world as opposed to the way that previous quantitative research has approached and analysed the Bank's evaluations at either 'scale' of aid (Mosley, 1986; Subramanian and Ragan, 2008; Picciotto, 2009; Arndt, Jones, and Tarp, 2010; Howes et al., 2011; Tierney et al., 2011).

Chapter Eight

Data, deceptions, and conclusions

8.1. Introduction

This dissertation has taken a critical approach to the analysis of development and aid effectiveness, exploring not 'if' it works but, rather, what the language and logic of the World Bank's evaluations reveal about the goals of development and how development efforts are measured. In so doing, I have worked towards reshaping the debate about 'aid effectiveness' by analysing the institutional context in and processes by which the data that inform these concepts are produced. The overarching conclusions concern how the institutional role of evaluation, the professional practice of evaluation, and the evaluation processes in the World Bank influence a particular narrative about 'aid effectiveness.' These findings clarify several issues about the meaning of key terms in international development, such as 'success,' and the evaluative practices that define them. This analysis, in my view, answers the three main research questions posed in the introduction to this dissertation. First, what is the context in which evaluation takes place at the World Bank and how does the role of evaluation within the Bank influence its evaluation reports? Secondly, what are the evaluation processes at the Bank, how do they function at different scales, and what do evaluations measure? Finally, who are the elite actors in the process of evaluation and how do they engage with and influence the results of evaluations in the Bank?

The answers to these questions and the conclusions of this dissertation as an entity are presented in this chapter. In presenting the conclusions, I draw on the findings from

the analysis of the processes of professional evaluation, the role of evaluators, and the institutional goals for evaluation. The answers highlight how these forces influence which data are included in evaluations, how determinations are made, and the logic by which 'success' or 'aid effectiveness' is defined in practice. I articulate these conclusions by bringing together and comparing the findings from all four empirical chapters. The conclusions suggest several key implications of the research for geographers and other academic researchers, scholars of development and evaluation policy, and for policymakers who use these ratings data to determine the impact of aid. There are, however, issues which are beyond the scope of this research. In the end, I present several available avenues for academic geographers and other scholars and for practitioners of international development to further develop the debate on 'aid effectiveness' in a way that considers the meaning of the data in context and that includes crucial perspectives in defining 'success' that are currently omitted from the Bank's evaluation processes.

8.2. A conclusion in three parts

There are three main statements that constitute the conclusion about how evaluation works in the Bank and the institutional factors that shape the 'window' through which evaluative data is produced. The key parts of this conclusion are explained and justified in the subsequent sections.

Data and the goal and role of evaluation (8.2.1.): I find that the data upon which 'success' is determined in the Bank are shaped by IEG's institutional role and purpose, primarily to determine the extent of the Bank's accountability. The data are also shaped by the role of evaluation in the Bank. This role is independent from the rest of the World Bank, including the operations teams that implement the Bank's projects and

programmes. As a result of performing this institutional role and function, the Bank's evaluation professionals use limited types of quantitative data which reflect only specific aspects of the promises and results of development rather than the myriad of issues that, based on the feedback from the evaluators themselves, should be considered in order to understand the impact of aid and what makes a project or programme 'successful.'

The evaluation process and the meaning of the ratings (8.2.2.): I argue that the evaluation process serves a 'regulative' role in the Bank in that it 'makes the rules,' in this case, the Bank's 'ratings' and other judgments. The reason for a successful or unsuccessful rating in the Bank's evaluations frequently reflects the institutional goals of evaluation, the requirements of the evaluation process at different 'scales,' and the strength of the data rather than an assessment of the impact of the Bank's work within a particular context.

'Success' and attribution in context (8.2.3.): Contrary to the assumption of previous studies of aid effectiveness (Mosley, 1986), measuring 'success' at the project- and country-level includes many variables that are not directly related to economic growth, at least as it is defined in the Bank's evaluations. Evaluators reported that the impacts of the Bank's projects would also not necessarily, in sum, lead to economic growth within a recipient country. I argue that, given the processes by which aid ratings are created in the Bank and the difficulty of demonstrating attribution in practice, it makes sense that the Bank's aggregate determinations of individual projects will not necessarily 'add up' to economic growth on the country-level.

8.2.1. Data and the goal and role of evaluation

In order to explain the production of the results of aid in the Bank, it is important to clarify what data and information the Bank's evaluations are based upon. Perhaps the most unexpected conclusions concern the role and goals of evaluation and how these factors shape the kinds of data upon which the Bank's evaluators base their reviews of the Bank's projects and programmes. The connection between the kinds of information used and the role of evaluation within the Bank is an important finding, which I confirmed in both the interviews and content analyses. The types and nature of the data that evaluators use in the evaluations reflect the institutional context in which evaluators work and the limited scope of the aspects of development which are considered.

The first part of this conclusion concerns the role of evaluation and how the Bank's institutional goals of evaluation influence the data used in this professional practice, themes which were developed in chapters four and five. First, evaluation is independent in order for evaluators to provide 'objective' determinations. Due to evaluation's independent role in the Bank, evaluators are not allowed to design monitoring and evaluation systems, collect data, or to interact with the Bank's projects and programmes until they are completed. The operations teams that create and provide this data are judged by the results of these evaluations. The institutional function of evaluation is important to consider: the Independent Evaluation Group (IEG) is responsible for determining, as objectively as possible, the extent of the Bank's accountability to its clients, or the extent to which the Bank 'did what it said it would do' (Grasso et al., 2003). As evaluators' responses in the interviews reflected, there are ways in which this institutional space limits what they are able to consider about projects. The ultimate evaluations are a snapshot of the extent to which the Bank completed its promised

deliverables and the extent to which they can be quantified. From the evaluators' perspective, these evaluations do not, for example, generally provide insight into the quality of decisions made during the development process or suggest ways in which the Bank's projects or programmes may have long-term impacts on economic growth or other overarching goals.

There are important links, therefore, between the 'evaluation space' in the Bank and the data that evaluators use in these evaluations. In the institutional context of the Bank, the practice of evaluation revolves around reviewing data and reporting the deliverables as defined in the original development objective rather than comprehensively considering the Bank's impact in the context of specific countries. There are many factors that evaluators consider within the analyses, as I have discussed in the analyses of the interviews with evaluators and the written evaluations. This appears to stand in contrast to how the Bank's evaluation results are generally conceived as a straightforward measure of the 'success' of development overall and that these project-level results, in sum, are a reflection on the status of 'aid effectiveness' (Mosley, 1986). In reviewing the Bank's evaluations, it is important to consider that evaluators' engagement with the process of development is often limited to their review of the data provided to them, not on their first-hand knowledge of the Bank's work or their own observations of its impacts in the recipient countries.

The second part of this conclusion relates to the characteristics of the quantitative data upon which evaluations are based and how the data are affected by the institutional role of evaluation and the authority of evaluators. The types and the nature of the data upon which the Bank's evaluations are based, the analysis of which was presented in chapters six and seven, reflects how evaluators described their role and the goals of

evaluation. There are two important aspects of this quantitative data that suggest the range of what these evaluations consider about aid. To begin with, these quantitative data points are collected to prove the Bank's accountability in meeting its promises to the client. They measure the deliverables that reflect the extent of the progress towards the Bank's development objectives and are not necessarily indicative of the Bank's contribution to economic growth in the recipient state. The data in the evaluations focus specifically on the deliverables that can be counted and quantified. The majority of the data in the Bank's evaluations reflects, for example, the number of outputs and the number of pieces of legislative data passed rather than the quality of these development outputs or why they are important to serving particular ends in the client country. In the written evaluations, evaluators do not investigate the quality of these outputs or their relevance for the host community's or the country's overall well-being and progress, economic or otherwise. In reviewing the evaluation documents themselves, I found that these evaluations do not include or strongly value the opinions of the beneficiaries or clients and generally focus on the factors that can be easily quantified rather than issues such as empowerment or social trust, which, evaluators suggested, are more difficult issues to assign numerical values. This generally confirms the previous literature on auditing professionals within institutions (Strathern, 2000).

Evaluators' views of their work within the Bank expressed during the interviews reflected many aspects of the data used in the evaluation reports that were analysed in chapters six and seven in this dissertation, especially the opinion that evaluation is an objective activity that should focus on quantitative analysis. It is important to highlight that this quantitative data are, in the evaluations I reviewed, neither place specific nor put into context of the environment or communities in which they were collected. This

approach reflects how evaluators described the evaluation process in the interviews: The Bank evaluators described evaluation as a universal knowledge that can be determined through common methods rather than explaining evaluation as a situated knowledge that depends on the context, culture, and history of the places in which development happens. Yet, in the country-level evaluations, it is important to note that there is some attention to the historical and developmental context of the country built into the evaluations, which sometimes contradicts how evaluators describe the process.

The quantitative nature of the data reflects the skills and goals that evaluators identified for their job within the Bank. While evaluators claimed they need a 'broad church' of skills to do their jobs, they identify themselves and the importance of evaluation in terms of their skill-set in economics, which is primarily related to quantitative abilities. In defining this unique skill set, evaluators often characterised evaluation as an 'art,' not a science. This phrase was used to assert that there is a certain skill-set that evaluators have which allows them to make accurate determinations that other staff within the Bank are not in the position to make. In effect, evaluators perceive that only they have the objective eye to understand the development process that other Bank staff cannot have due to their position in the implementation process.

8.2.2. The evaluation process and the meaning of the ratings

Another important contribution of this dissertation has been to identify and analyse the issues that are most important to evaluators when they determine the ratings for the Bank's projects and programmes. Understanding the details of the process of these decisions and what distinguishes 'successful' and 'unsuccessful' ratings is important because the ratings are used in studies on aid effectiveness. The previous

studies on aid effectiveness do not consider what these ratings reflect, basing their analysis only on the calculated rate of return for projects in a country compared to the economic growth in the country (Mosley, 1986). In my estimation, the analysis presented in this dissertation clarifies what these ratings actually reflect about the Bank's work and reveals the limitations of using the ratings data in quantitative analyses without considering the institutional realities which shape it.

The power of evaluators' decision-making power within this evaluation process is another conclusion that demonstrates how the role of evaluation within the Bank shapes the final results. I found that the Independent Evaluation Group (IEG) is authoritative within the Bank and not only determines these ratings, but also 'produces' them and makes the authoritative decisions in determining the overall extent of the Bank's accountability. In this, the evaluation team has a 'regulative role' in that it is able to 'make the rules' (Scott, 2001). Understanding the evaluators' viewpoints and the role of these professionals is, therefore, critical to assessing the meaning of the ratings. Within this development space, the professional practice of evaluation and the viewpoints of the evaluators, who hold influential roles within the organisation, have the authority within the Bank to select which types of these data are valued and seen as most critical for projects and programmes to prove in order to be successful. In chapter five, I detailed how evaluators identified several aspects of the evaluation process itself that influence the ratings that have less to do with the impact of a project on a country's economic development and more to do with the evaluation process as it currently stands. An additional institutional influence that effects these determinations is how evaluators view themselves and their work, as I discussed in chapter four. In speaking to evaluators, I found that they explained their main skills as being related to their backgrounds in

economics and their quantitative abilities. In the context of the Bank's evaluation process, the ratings are most likely a reflection of the strength of the data that are collected rather than the impact of the project. Projects that present strong outputs will not necessarily be strongly rated unless they have a systematic monitoring and evaluation system in place and vice versa.

The process of identifying the factors that 'tip the balance' between successful and unsuccessful projects and programmes in the Bank's evaluations helped to elucidate the thought process of the Bank's evaluators. In comparing the interviews and the content analyses, I found that, contrary to the assumptions of previous studies, the project- and country-level ratings are based upon many factors that are not related to the Bank's contribution towards economic growth and, interestingly, reflect many of the limitations that evaluators identified with the mandates of the Bank's evaluation processes at both the project- and country-level. In the written evaluations at the project level, evaluators tend to focus on the perception of the operations staff and the quality of the data presented rather than analysing the impact of the data that are provided in the self-evaluation written by the Bank's operations staff. At the country-level, the Bank's evaluators quantitative data about the Bank's deliverables rather than discussing the specifics of the Bank's work that led to particular outcomes. Evaluators do not, however, investigate reasons for the Bank's impact or the meaning of these quantifiable indicators for the country and its people, which other models would (Pawson, 2003).

The lens of 'scale' in terms of the evaluation process at the Bank has been critical to the analysis throughout this dissertation. Rather than being one evaluation process, there are several processes within the Bank that judge the impact of the Bank's work at different 'scales,' or the geographical scope of the Bank's work. I focused on the project-

and country-level reports in particular. There are different institutional goals and processes of evaluation at the project and country-level which evaluators identified in their interviews. Given what evaluators say about project-level evaluations and the data presented in these evaluations, it is clear that they are more of a 'validation' than an investigation into the Bank's role and impact. In the interviews, evaluators described the country-level evaluations as being more in-depth and requiring more in-country investigation and more innovative data collection by evaluators than the project-level evaluations. In the content analysis, I found that the country-level evaluations do include more interaction with the country and more data collection and that the project-level evaluations are, as evaluators suggested, much more limited and formulaic in their approach and considerations. Overall, this suggests that there are multiple meanings of aid effectiveness, depending on the context in which these determinations are construed.

There is, however, one critical difference between what I found in the project- and country-level evaluation processes and how evaluators described them to me in the interviews. Although evaluators explain the practice of evaluation differently in terms of the goals and uses of the practice at different scales in the World Bank, at both scales of evaluation, the actual evaluation reports reflect a similar methodology: evaluators generally take similar steps, focusing on the deliverables and the quantity of the data rather than assessing the relevance of that data. In evaluations that included in the samples at both scales, evaluators present indicators and project results to show the sum of the Bank's work quantitatively. The current framework does not require evaluators to analyse the link between their project-level and country-level results. This finding points to a commonality in the limited scope of the data and the processes that evaluation at the Bank currently employs at both scales of aid.

As well as understanding the methodology of evaluation, it is important to consider the language by which evaluators justify 'success' and the context in which certain terms are used (Fairclough, 2013), as this is an important part of what defines the process. This language reflects the limitations of the role of evaluation, which evaluators identified in the interviews, and how this role impacts the evaluations and ratings in demonstrable ways. At both scales of evaluation, evaluators are inclined to use language that relates to the Bank's evaluation standards to justify their determinations. Although it is not unexpected that the Bank's evaluators reference these evaluation standards, the way in which evaluators make use of these standards is, somewhat, unanticipated: evaluators often reference these evaluation standards in lieu of completing an in-depth review of the impact of the Bank's work. In effect, the language that evaluators used in the evaluations confirms that the goals of these documents is to prove the Bank's accountability in an objective and codified way rather than thinking about the impact of the development outputs. At the project-level, the determining factor that the evaluators consider in assigning a particular rating has to do with the mandates of the Harmonized Criteria rather than an overarching assessment of the project. At the country-level, rather than considering the contribution of the project to the country in historical context or even the strength of the programme based on the elements they reviewed, evaluators often sum the results of these projects within areas such as health and education projects to determine the Bank's overall impact for the relevant pillar within the country. In using the data created in these processes, therefore, one must recognise this process and that the assessment may be due to a mathematical weighting and bureaucratic considerations rather than a comprehensive assessment of the impact of the project for a particular community or country.

8.2.3. ‘Success’ and attribution in context

The final conclusion serves to clarify an important concept in context. One of the main questions from the start of this research was about the relationship between the practices of evaluation and the meaning of the concept of ‘aid effectiveness’ and ‘success’ in the context of the IEG’s evaluations. This builds, in part, from the previous finding about the differences and similarities in evaluation processes at different ‘scales’ of aid at the Bank. Following other studies, I maintain that there is a lack of theoretical clarity about what the term ‘aid effectiveness’ means as well as confusion about what it means in practice because it is “treated as common knowledge” (Hyden, 2008, 264). In the critical literature, ‘aid effectiveness’ is used to connote both a standard of the outputs of aid and the overall judgment of the international development system. I have limited the discussion of ‘aid effectiveness’ to how ‘success’ is defined in the Bank’s evaluations because these ratings constitute the data upon which larger claims related to international aid as a system are based. In considering ‘success’ and how it is defined in the Bank’s evaluations, I analysed the related issue of attribution and how the Bank evaluators develop this in practice. By dissecting the meaning of the key concept of ‘aid effectiveness,’ the analysis has engaged with and, in some ways, challenged the ‘micro-macro paradox’ theory, which assumes that there is a link between the percentage of the Bank’s ‘successful’ project-level ratings in a country and its economic growth (Mosley, 1986).

There are two main findings about the meaning of the concept of ‘aid effectiveness’ in context, which are related to the processes of evaluation within the Bank and how ‘success’ is codified and justified by evaluators in their written evaluations. First, the definition of aid effectiveness in practice within the Bank is most likely not simply a

reflection of the progress towards economic growth, as the 'aid effectiveness literature' implicitly assumes, and the information upon which 'success' is determined is far less about the economic rate of return than previous studies in the aid effectiveness debate have suggested (Mosley, 1986). Rather, in the Bank's evaluations, 'success' is defined in terms of many aspects of a project or a program, including: long-term development goals for the Bank in that area, project planning which connects to country-level development goals, projects which consider the stability of their environment, Bank staff that plans for challenges and responds to them, and in-country clients that are committed to implementing projects in the way the Bank believes to be most efficient. This conclusion results because the Bank's evaluations are an accountability tool, determining if the Bank did what it said it would do, rather than if the project under evaluation contributed towards a country's economic growth.

Secondly, contrary to the assumptions of the micro-macro paradox (Mosley, 1986), the Bank's aggregate determinations of individual projects within a country will not necessarily 'add up' due to what data are considered in project-level evaluations and the difficulty of attribution in practice. While the logic by which 'aid effectiveness' is determined on the two 'scales' that were analysed (the project- and country-level) is generally similar, evaluators in the interviews and the way in which Bank's evaluation reports describe attribution make it clear that it is difficult to prove attribution and that it would make sense that project-level results would not 'add up' to country-level economic growth. This is, in part, the result of the institutional role of evaluation in the Bank, as the Bank's evaluations currently use primarily quantitative indicators to determine whether it met its goals and the overall impact in the country. This institutional role, by extension,

makes it difficult for evaluators to complete the kind of detailed evaluation that the theory indicates would be required to determine attribution.

There are several obstacles to proving attribution which are reflected in how evaluators described the evaluation process in interviews and how evaluators discuss the issue in the written evaluations included in the content analysis. First, as evaluators mentioned during the interviews, there are often other initiatives being simultaneously implemented by the client states and other organisations. In several evaluations, evaluators note that it is difficult to ascribe the outcomes reported to the Bank's project as opposed to the other initiatives or the sum of these initiatives. Secondly, in the interviews, several evaluators referred to the steps between the programme design and the results of the projects as the 'black box' because they are not involved in the implementation of the project and, therefore, must rely on the information provided to them about the project's planning and its results rather than being able to see the project as it evolves over time. This also shows that, in practice, it may be difficult to prove the 'counterfactual' in evaluation in the Bank, or what would have happened in the absence of the Bank's project or programme. Evaluators identified in their written evaluations that was often not possible to prove the 'counterfactual' because of a lack of baseline data from operations against which the project's outcomes may be compared.

8.3. Implications

Given the interdisciplinary nature of the research, it is important to clarify both the audiences for the findings and how the findings may be useful to them. The overarching conclusion is relevant to the theoretical debate on aid and has practical applications for aid practitioners and researchers who work to improve the aid system, often relying on

the results of the Bank's evaluations to inform their understanding of particular development projects and programmes. The overall conclusions concern the function of evaluation within the Bank, the role of evaluators, and the meaning of the data contained in the Bank's evaluation reports as well as its limitations.

8.3.1. Implications for the multi-disciplinary academic literature

The dissertation contributes to the literature on monitoring and evaluation, to the geographical literature and other theoretical analyses of development, and to the theoretical development of the concept of 'aid effectiveness' which is important in several academic fields. First, by discussing the data and the process of making the determination about the success of particular efforts rather than focusing on the question of 'if aid works,' the analysis reflects an articulation of what the Bank's evaluative data do and do not reflect about the Bank's impact and what the role of evaluation is in practice. In analysing the practice of evaluation in context, I contribute to the understanding of what issues such as 'attribution' and 'logical frame' mean in the Bank and the institutional realities in which they have developed (Iverson, 2003; Rebien, 1996). By considering the issue of attribution and the similarities and differences in evaluation at the project and country-level, I provide insight into how aid and evaluation results exist at different 'scales,' which is an issue that previous analysis has not reviewed, essentially using the data contained in all evaluations as one group of knowledge and data. The analysis contributes to the theory of monitoring and evaluation by providing insight into how data are used in the logical framework of the Bank's project-level evaluations and how this log frame limits the scope of its evaluations to only certain aspects of the projects and programmes that constitute international development. In addition, I have analysed development and evaluation both as practices,

with specific and regimented actions, and as narratives, with certain processes that are intended to reflect the impact of aid to the world.

Next, the analysis in this dissertation is complementary to previous work in the fields of development geography and economic geography and particularly contributes to the current literature in development geography (Lawson, 2007), the post-structural analyses of aid (Escobar, 1995; Crush, 1995), and to the critique of the World Bank's approach to scale and place that began with its 2009 World Development Report (Peck and Sheppard, 2010; Harvey, 2009; Wilson, 2011). In this dissertation, I put forth a logical methodology by which geographers may use their unique lens, which is cognisant of issues such as 'scale' and 'narrative,' to engage with development organisations in their research in productive ways. The previous geographical analysis of international development has primarily focused on the 'little d' processes of development (Hart, 2001) and is highly critical of organisations such as the Bank and the impact they have on recipient communities due to the Bank's lack of context in its analysis and claims of universal knowledge (Lawson, 2007; Peet and Hartwick, 2009). This focus, however, means that there are relatively few studies that use the unique theoretical lens of geography to explore *how* the Bank and other development organisation work and how the assumptions imbedded in these institutions impact their public narrative. The study of the processes and elite actors that shape evaluation in this dissertation demonstrates how, from my position, the post-structural analysis of development (Escobar, 1995; Crush, 1995) can be useful to demonstrating the specific ways in which language and logic are used within evaluation teams in powerful institutions. For example, I have identified important ways in which the role of evaluation within the Bank influences the elite actors' engagement with the places of development, providing context for the

complementary claims of economic geographers who are critical of how the Bank conceives of the spaces of development and works with its clients in these spaces (Harvey, 2009).

Thirdly, these conclusions challenge the previous literature in development economics on aid effectiveness, which has claimed there is a 'paradox' between the successful results of the Bank's projects and the economic growth of its client countries. The literature on the micro-macro paradox (Mosley, 1986; Subramanian and Ragan, 2008) is based on two assumptions: first, that a majority of 'successful' projects in a country would lead to economic changes in the recipient country and, secondly, that there is a clear relationship between the results of the Bank's projects and country-level outcomes. I have investigated both how the Bank assesses its impact at different levels and the links between project and country-level impacts and have found specific ways that the evaluation process, rather than the impacts of aid itself, affects the perception of how well aid 'works.' By considering more closely what the Bank's evaluations reflect about the project impacts and what they do not, the methodology and the assumptions upon which this theory is based have been challenged. Of particular consequence to the micro-macro paradox theory, the Bank's ratings draw on a wide array of data, which often are not related to economic return, and that there is not necessarily a link between project-level success and economic growth on the country-level, in part due to the difficulty of determining attribution in a complicated environment such as the countries in which the Bank works.

This dissertation has additional important implications for literature which considers recent international efforts to improve aid. In particular, the analysis about what evaluators consider in the process of completing evaluations and which issues are

most important in the determinations at the Bank elucidates the most important issues by which 'success' is being defined and judged. There is an expanding view of what aid should do to 'work,' which now includes a multidimensional view of both poverty and the goals for aid (Sen, 1999) and is codified in efforts such as the Poverty Reduction Strategy Papers approach and the Millennium Goals (Mawdsley, 2007). Given the importance of the declarations on aid effectiveness and efforts to achieve ends such as alleviating poverty through aid, I maintain that it is critical to understand the extent to which these declarations are simply aspirations, or if they have served to actually change how the practice of international development operates. The extent to which influential organisations such as the Bank have integrated these adjusted expectations into their practices and the rubrics by which they judge their own work is important to understand. In this dissertation, I have explained the way that the Bank currently considers issues such as poverty alleviation and the education of women and girls, but only the aspects which can be strictly quantified in its overall determinations; this data does not recognise many aspects of the impact of the Bank's work which may be important.

8.3.2. Implications for development policy and evaluation systems

In presenting these dissertation findings, I have clarified the role of evaluation in the Bank from the perspective of those engaged in it and have identified some of the limitations in the evaluation process in the Bank as it currently stands. In so doing, I contribute to the understanding of the professional practice of evaluation, how it operates, and the factors that influence how the Bank evaluates its work and reports its findings. The audience that will most likely find the implications of this dissertation useful are those who use the information that the Bank's evaluations provide in a decision-making

capacity. Client governments and other development organisations, for example, use the Bank's evaluative data to decide which programmes to implement. One of the implications of this research is that governments and development organisations can make better decisions by understanding the source and the limitations of this data. By being more aware and informed about the 'window' through which evaluative data and determinations are created and influenced, these practical decisions about programme design and the theoretical claims about the impact of aid can better match the reality of these processes on the ground.

This research also has implications for those working in operations and other areas of the Bank and provides insight that can help the professionals in these areas of the Bank work together more productively. One of the issues this study highlights is the possibly problematic nature of the relationship between the operations staff and evaluators in the Bank which is, in large part, due to the fact that the determinations from the Bank's evaluations impact the career progressions of operations staff. An improved understanding of the ratings process through the eyes of the professionals that carry out this process could help to make the process more transparent to those who are judged by it and illuminate the misconceptions that I noted among the operations staff in the Evaluation Week Conference in 2013 (Yannias, 2013a). In addition, there are implications for Bank staff members currently employed in impact evaluation, who may use the findings of this dissertation to better understand the institutional goals for IEG and how it works. Using the conclusions of this dissertation, the Bank's leadership and its staff may consider how the evolving process of impact evaluation can complement the ongoing evaluative processes in the Bank, in part by focusing more on data about attribution and the Bank's impact for particular projects and programmes.

Finally, there are important policy implications for those working to design or reform evaluation systems in other international development organisations and agencies. The fact that the Bank's determinations are based on reporting quantitative data rather than a comprehensive analysis of the relevancy of this data suggests how the chosen role of evaluation within an institution controls how the data are collected and the processes by which it is reviewed. I have particularly highlighted the implications of having evaluation as an 'independent' force within a larger institution and how this limits what evaluations are able to convey about the impact of the organisation's work. With these findings in mind, those designing or reforming other systems can, perhaps, design alternative institutional contexts for evaluation that will be able to assess the wider range of variables that are important in the context of aid projects and programmes.

I have particularly investigated the implications of different modes of evaluation that the Bank is currently pursuing, impact evaluation. Since evaluation is a changing field, this dissertation has implications for organisations seeking to use impact evaluation as a methodology or to introduce it as a new evaluation unit (Banerjee et al., 2007). As the Bank continues to invest heavily in impact evaluations (World Bank DIME, n.d.), it is important for other development organisations and the Bank leadership to consider carefully the existing state of evaluation knowledge within the Bank and how the emergence of newer departments such as the new impact evaluation unit (DIME) affect the overall evaluation space in the Bank.

8.4. Directions for future research

In this dissertation, I have focused on roles, processes, and elite actors involved in evaluation at the Bank. Due to the decisions made in the research process, there are

several important issues and considerations about how the Bank works that were beyond the scope of this dissertation. There are three specific research areas that would help in particular to further the understanding of aid effectiveness and the process of evaluation: a comparison of evaluations completed in different geographical locations and over time, an investigation of the beneficiaries' views of the Bank's evaluative practices, and research concerning how the Bank's leadership uses evaluative data in its planning process and in financial decisions.

Here, the focus has been on the evaluation process itself rather than a particular geographic area or type of development project. This dissertation highlights the need for further research, particularly qualitatively-based research, to explain the meaning and implication of the Bank's ratings in the context in which they are 'produced,' drawing on the approach used in post-structural studies (Escobar, 1995; Crush, 1995). An important issue for future research and for those who continue to analyse the narrative of development is to consider how the Bank's ratings compare to each other in different contexts, using different variables such as geographic areas, project categories, and time. In this way, academic researchers, especially those in the subfields of development geography and economic geography (Lawson, 2007) that have criticised the Bank for its lack of attention to space and place could make an important contribution to understanding how evaluation, as a particular professional practice, produces certain explanations of these spaces and the impact of the professional practices of international development on them. The content analysis of agricultural projects in chapter six, for example, highlighted the considerations in and issues with the Bank's project-level evaluations, but does not consider the different issues that the Bank possibly considers in different types of projects. The content analysis of African country-level evaluations in

chapter seven does not provide a comparative view of how the Bank's evaluators measure the progress of programmes in different African countries or different continents.

The choice to focus on the professional practice of evaluation and the institutional role of this practice means that the interviews were with evaluation professionals rather than the beneficiaries of aid projects. The analysis of the current state of evaluation suggests that the Bank's evaluators do not engage in issues relevant to the places of development for the most part and do not include the viewpoints of beneficiaries in their evaluation reports. The perspective and the opinions of beneficiaries are essential in the evaluative process, as critics of the Bank's work have emphasised (Chouinard, 2013). An important area of research, therefore, would be to compare the goals of the recipients of aid projects and programmes to the issues actually considered in the Bank's evaluations of these projects and programmes. By comparing and contrasting the current evaluation rubric in the Bank with the goals articulated by those who are the recipients of these programmes, it may be possible to identify important issues that the Bank's current evaluations do not consider and that will more comprehensively explain the impact of efforts in international development.

Finally, because the interviewees for this project were evaluators rather than the leaders in the Bank who make financial and policy decisions in the Bank, one of the issues that this dissertation has not addressed relates to how evaluative knowledge is used in other areas of the Bank to design new programmes and to determine financial disbursements. In the interviews, evaluators reported that they were often unsure about how evaluation knowledge is utilised in the Bank and how evaluative information informs funding and other decisions within the Bank. Additionally, an issue that would

be useful to research is the decision-making process by which the Bank's leadership decides to complete country-level evaluations and why the Bank completes comprehensive evaluations in some countries and not in others. From reviewing the publically available data, the Bank has completed more country-level evaluations in the more economically affluent countries: in fact, the three countries with the most World Bank country-level evaluations (Brazil, India, and Indonesia) are the seventh, tenth, and sixteenth countries in terms of GDP (World Bank, GDP Rankings). By analysing why the Bank completes country-level evaluations in some countries but not in others, and how it uses the information that it does collect, researchers could perhaps better explain the multiple forces that shape the scope of evaluation data about the Bank's country-level impacts.

These avenues for future research suggest a major research agenda, requiring some changes in how academics and policymakers engage with the practice of evaluation in international development and conceptualise the term 'aid effectiveness.' Of particular relevance to the subfields of development and economic geography, they would require more engagement with the specific practices within the Bank and the 'big D' development processes of which they are a part. This dissertation is, therefore, one step in a wider effort to help those using the Bank's evaluative data for policymaking and academic research to make better-informed decisions by analysing the details and the implications of the evaluation data created in the influential halls of the World Bank.

Appendices

Appendix 1: Summary of Interviews

	Description	Current place of employment	Interview month
1.	Academic expert on aid agencies with consulting experience in an agency	University	February 2013
2.	Academic expert on international development theory and former consultant	University	February 2013
3.	Academic expert with operational and evaluation experience in development	University	February 2013
4.	Academic expert on monitoring and evaluation with experience conducting research in an aid agency	University	February 2013
5.	Academic expert on international development with experience conducting research in an aid agency	University	February 2013
6.	Current World Bank Evaluator	Independent Evaluation Group (IEG), World Bank	March 2013
7.	Current World Bank Evaluator, more than 10 years experience	IEG, World Bank	March 2013
8.	Current World Bank Impact Evaluator	Development Impact Evaluation Initiative (DIME), World Bank	March 2013
9.	Current World Bank Impact Evaluator	DIME, World Bank	March 2013
10.	Current World Bank Impact Evaluator	DIME, World Bank	March 2013
11.	Current World Bank Evaluator	IEG, World Bank	March 2013
12.	Current World Bank Evaluator	World Bank, but not IEG	April 2013
13.	Current World Bank Evaluator	World Bank, but not IEG	April 2013
14.	Current World Bank Evaluator, more than 10 years experience	IEG, World Bank	April 2013
15.	Current World Bank Evaluator	World Bank, but not IEG	April 2013
16.	Current World Bank Evaluator	IEG, World Bank	April 2013
17.	Current World Bank Evaluator, more than 10 years experience	IEG, World Bank	May 2013
18.	Current World Bank Evaluator	IEG, World Bank	September 2013
19.	Current World Bank Evaluator, more than 10 years experience	IEG, World Bank	September 2013
20.	Current World Bank Evaluator, more than 10 years experience	IEG, World Bank	September 2013
21.	Current World Bank Evaluator	IEG, World Bank	September 2013
22.	Former World Bank Evaluator, more than 10 years experience	Evaluation team at different aid agency	September 2013
23.	Current World Bank Evaluator	IEG, World Bank	September 2013
24.	Current World Bank Evaluator	IEG, World Bank	September 2013
25.	Current World Bank Evaluator	IEG, World Bank	September 2013

Appendix 2: Interview Questions

Introduction

- Review my research and address any initial concerns about confidentiality

Understanding your job

- How did you decide to get into evaluation?
- How would you describe the key skills that someone needs to do your job?
- What have you learned about foreign aid through this kind of work?
- What do you think are some common misconceptions about evaluating foreign aid?

Data and Measurement

- How would you describe the evaluation process? What do you see as the goal and importance of evaluation?
- What are the key things you look at when you review an Implementation Completion Report (ICR) written by someone in operations?
- How does information from the ICRs get used in country reports? Which information in the self-assessments is most important in your opinion?
- Who do you go to for technical assistance if you need it?
- What is not being measured at the project- or country-level that should be measured?
- What cannot be measured through the current methods of evaluation that should be included in the project evaluations in your view?
- What is the most significant change you have seen in evaluation methodology or in terms of its organisational role?

Role of evaluation in World Bank

- How does the World Bank decide to do a country report on a particular country? How are the years of the review chosen and who makes these decisions?
- Who do you collaborate with here at your organisation? Do you generally only work with other professionals who do aid evaluation or is there a lot of cross-departmental collaboration?
- What does the World Bank do with the information that is created in evaluations? Who reads these reports and how is this used for policy making?

Appendix 3: Project-level evaluations (Implementation Completion Report Reviews) included in sample for content analysis

	Project name	Location	ICR Review date
1.	Avian Influenza Preparedness Project	Bosnia and Herzegovina	12/2013
2.	Integrated Management of Aquatic Resources in the Amazon	Brazil	12/2013
3.	Vietnam Water Resources Assistance	Vietnam	12/2013
4.	South Kordofan State Emergency Project	Sudan	12/2013
5.	Sustainable Agro-pastoral and Land Management promotion	Cameroon	12/2013
6.	Integrated Management of Critical Ecosystems Project	Rwanda	12/2013
7.	Preparedness Avian Influenza Control and Human Influenza Emergency Preparedness and Control	Haiti	11/2013
8.	Livestock and Fisheries Development Project	Sudan	11/2013
9.	Avian Influenza Control and Human Pandemic Preparedness and Response Project	Georgia	11/2013
10.	Caribbean Implementation of Adaption Measures in a Coastal Zone	Caribbean	11/2013
11.	The Greater Addo Elephant National Park Project	South Africa	11/2013
12.	Post-conflict Reconstruction and Recovery Grant	Cote d'Ivoire	11/2013
13.	Avian Influenza and Human Pandemic Preparedness and Response	Turkey	9/2013
14.	Andhra Pradesh Rural Poverty Reduction Project	India	9/2013
15.	Avian and Human Influenza Control and Preparedness Project	Laos	9/2013
16.	Community-based Rural Land Development Project	Malawi	8/2013
17.	Car Food Response Project	Central African Republic	8/2013
18.	Avian and Human Influenza Control and Prevention	Egypt	6/2013
19.	Avian Influenza Control and Human Pandemic Preparedness and Response Project	Romania	6/2013

20.	Rural Development Project	Georgia	6/2013
21.	Economic Management and Governance Reform Grant	Central African Republic	6/2013
22.	Agricultural Development and Credit Project	Azerbaijan	6/2013
23.	Livestock Waste Management in East Asia	East Asia and Pacific	6/2013
24.	Cn-heilongjiang Diary	China	6/2013
25.	Avian Influenza Control and Human Pandemic Preparedness and Response Project	Kosovo	6/2013
26.	Land Administration Project	Ghana	6/2013
27.	Biodiversity Conservation and Sustainable Development Project	South Africa	6/2013
28.	Id National Program for Community Empowerment in Rural Areas	Indonesia	6/2013
29.	India: Emergency Tsunami Reconstruction Project	India	5/2013
30.	Micro, Small and Medium Enterprise Project	Nigeria	5/2013
31.	Hai Basin Integrated Water and Environment Management Project	China	5/2013
32.	Natural Resources Development Project	Albania	4/2013
33.	Africa Emergency Locust Project	Africa	4/2013
34.	Poverty Reduction Support Grant	Benin	4/2013
35.	Agricultural Technology Transfer Project	China	4/2013
36.	West Delta Water Conservation and Irrigation Rehabilitation Project	Egypt	1/2013
37.	Tg-economic Recovery and Gov Grant	Togo	1/2013
38.	Rural Investment and Local Governance Project	Cambodia	12/2012
39.	Support for Economic Expansion and Diversification	Zambia	12/2012
40.	Avian Influenza Control Project	Nepal	12/2012
41.	Rural Enterprise and Small-scale Commercial Agriculture Development Project	Armenia	12/2012
42.	Amapa Sustainable Communities	Brazil	11/2012
43.	Real Estate Cadastre and Registration Project	Serbia	11/2012
44.	Avian Influenza Control and Human Pandemic Preparedness and Response Project	Moldova	10/2012
45.	Forest Development and Conservation Project	Bosnia and Herzegovina	9/2012
46.	Ecosystem Restoration of Riparian Forests in Sao Paulo	Brazil	9/2012

47.	Avian Influenza Control and Human Pandemic Preparedness and Response Project	Nigeria	9/2012
48.	Integrated Community-based Ecosystem Management	Namibia	8/2012
49.	Avian Influenza and Human Pandemic Preparedness and Response	Albania	8/2012
50.	Economic Governance and Recovery Grant	Cote d'Ivoire	8/2012
51.	Social and Economic Recovery Project	Croatia	8/2012
52.	Second Community Nutrition Project	Madagascar	8/2012
53.	Agricultural Research and Extension APL	Peru	8/2012
54.	Tien Shan Ecosystem Development Project	Central Asia	7/2012
55.	Nuestras Raices Program	Honduras	5/2012
56.	Irrigation Distribution System and Management Improvement Project	Azerbaijan	4/2012
57.	Price Vulnerability (Food Crisis)	Nicaragua	4/2012
58.	Syr Darya Control and Northern Aral Sea Phase I Project	Kazakhstan	2/2012
59.	Economic Recovery and Gov Grant 3	Togo	10/2011
60.	Ethiopia Global Food Crisis Response	Ethiopia	10/2011
61.	Ci-ivc Pngter Rural Land Management	Cote d'Ivoire	8/2011
62.	Avian Influenza Preparedness Project	Armenia	8/2011
63.	Avian Influenza and Human Pandemic Preparedness and Response	Tajikistan	7/2011

Appendix 4: All sections of the ICR Review, by information provided and rating

Section number	Section name	Information contained	Official rating involved?
1.	Project Data	Project details, sectors, themes, basic financial information, and names of those involved in the project and evaluation	No
2.	Project Objectives and Components	a) Objectives b) Were the project objectives/key associated outcome targets revised during implementation? c) Components d) Comments on project cost, financing, borrower contribution, and dates	No
3.	Relevance of Objectives & Design	a) Relevance of objectives b) Relevance of design	Classified as 'modest' or 'substantial,' but this is not an official rating
4.	Achievement of Objectives (Efficacy)	Objectives listed followed by the outcomes (often in quantifiable terms)	Classified as 'modest' or 'substantial,' but this is not an official rating
5.	Efficiency	a) IEG's assessment of the project's outcomes listed in Section 4 b) If available, enter the Economic Rate of Return (ERR)/Financial Rate of Return at appraisal and the re-estimated value at evaluation (quantitative value)	Classified as 'modest' or 'substantial,' but this is not an official rating
6.	Outcome	An explanation of the extent to which the project achieved its objectives and a rating	Rating on the 6 point scale (Unsuccessful to Successful)
7.	Rationale for Risk to Development Outcome Rating	An explanation of the extent to which the IEG evaluator finds evidence in the ICR to support the sustainability of the project in the long run.	Rating on the 6 point scale (Unsuccessful to Successful)

8.	Assessment of Bank Performance	a) Quality at entry (assessment of the strength of the project design) b) Quality of supervision	Ratings for both sections on the 6 point scale (Unsuccessful to Successful)
9.	Assessment of Borrower Performance	a) Government performance b) Implementing agency performance	Ratings for both sections on the 6 point scale (Unsuccessful to Successful)
10.	M&E design, implementation, and utilization:	a) M&E design b) M&E implementation c) M&E utilization	One overall rating for M&E quality as Substantial/modest
11.	Other issues	a) Safeguards b) Fiduciary compliance c) Unintended impacts d) Other	None
12.	Ratings	Ratings for the four aspects of the report that receive ratings are presented (Outcome, Risk to Development Outcome, Bank Performance, and Borrower Performance). The Quality of the ICR is also assessed by IEG and included in this table.	The table has three columns: the ICR ratings (by operations), the ICR Review ratings (by IEG), and any reasons for disagreement and comments.
13.	Lessons	This identifies the lessons that IEG says are highlighted in the ICR in terms of project design, implementation, or any other factor.	None
14.	Assessment recommended	Yes or No answer	Yes or No answer
15.	Comments on the quality of the ICR	IEG assessment of if the operations staff's ICR is 'satisfactory' or 'unsatisfactory.'	Yes or No answer

Appendix 5: Publically available country-level evaluations in African countries

Number	Country	Type of evaluation	Publication year
1.	Liberia	Country Program Evaluation (CPE)	2012
2.	Mozambique	Country Program Evaluation (CPE)	2011
3.	Nigeria	Country Assistance Evaluation (CAE)	2010
4.	Uganda	Country Assistance Evaluation (CAE)	2009
5.	Mali	Country Assistance Evaluation (CAE)	2007
6.	Madagascar	Country Assistance Evaluation (CAE)	2006
7.	Malawi	Country Assistance Evaluation (CAE)	2006
8.	Senegal	Country Assistance Evaluation (CAE)	2006
9.	The Islamic Republic of Mauritania	Country Assistance Evaluation (CAE)	2005
10.	Kenya	Country Assistance Evaluation (CAE)	2004
11.	Zimbabwe	Country Assistance Evaluation (CAE)	2004
12.	Eritrea	Country Assistance Evaluation (CAE)	2004
13.	Rwanda	Country Assistance Evaluation (CAE)	2004
14.	Zambia	Country Assistance Evaluation (CAE)	2002
15.	Lesotho	A Joint World Bank–African Development Bank Evaluation	2002
16.	Uganda	Country Assistance Evaluation (CAE) and a participatory evaluation	2001
17.	Cameroon	Country Assistance Evaluation (CAE)	2001
18.	Ethiopia	Country Assistance Evaluation (CAE)	2000
19.	Togo	Country Assistance Evaluation	2000

20.	Kenya	Country Assistance Evaluation (CAE)	2000
21.	Malawi	Country Assistance Evaluation (CAE)	2000
22.	Tanzania	Country Assistance Evaluation and IFC Country Impact Review	2000
23.	Burkina Faso	Country Assistance Evaluation	2000
24.	Ghana	Country Assistance Evaluation	2000
25.	Ivory Coast (English)	Country Assistance Review	1999
26.	Mozambique	Country Assistance Review	1998
27.	Zambia	Country Assistance Review	1996
28.	Ghana	Country Assistance Review	1996

Appendix 6: Publically available Country Assistance Strategy Completion Report (CASCR) in African countries

Number	Country	Evaluation type	Year of review
1.	Nigeria	CASCR Review	2014
2.	South Africa	CASCR Review	2013
3.	Ghana	CASCR Review	2013
4.	Guinea	CASCR Review	2013
5.	Mauritania	CASCR Review	2013
6.	Liberia	Joint Country Assistance Strategy (JCAS)	2013
7.	Democratic Republic of the Congo	CASCR Review	2013
8.	Niger	CASCR Review	2013
9.	Benin	CASCR Review	2013
10.	The Gambia	JASCR	2013
11.	Malawi	CASCR Review	2013
12.	Zambia	CASCR Review	2013
13.	Senegal	CASCR Review	2013
14.	The Republic of Congo	CASCR Review	2012
15.	Burundi	CASCR Review	2012
16.	Ethiopia	CASCR Review	2012
17.	Mozambique	CASCR Review	2012
Not available: Burkina Faso			

Appendix 7: Pillar Themes in Country Program Evaluation (CPE), Country Assistance Evaluation (CAE), and Country Assistance Strategy Completion Report (CASCR) Documents

Evaluations which include this as a pillar (related goals indicated in brackets)		
Pillar	**Country Program Evaluation (CPE)/ Country Assistance Evaluation (CAE)**	**Country Assistance Strategy Completion Report (CASCR)**
Growth		
Pro poor growth	1. Liberia 2. Burkina Faso (economic growth and poverty alleviation)	1. Mauritania (anchor growth in the economic environment of the poor) 2. Liberia
Macroeconomic stabilization	1. Mozambique (2011) 2. Uganda (growth) 3. Nigeria (governance) 4. Malawi (2006) (public sector reform) 5. Malawi (2000) 6. Senegal (sustainability, private sector development, infrastructure, agriculture and rural development) 7. Zimbabwe (promoting growth) 8. Zambia (providing balance of payments support to overcome severe external imbalances) 9. Ethiopia (maintenance of macroeconomic stability) 10. Kenya (2000) 11. Tanzania (improving the quality of fiscal adjustment in IMF-led stabilization) 12. Ghana (restore and maintain macroeconomic stability) 13. Togo	1. Mauritania (accelerate growth and maintain macroeconomic equilibrium) 2. Zambia (macroeconomic and expenditure management)

Financial reforms	1. Ethiopia (competitive and flexible exchange rates, liberalization of foreign trade) 2. Ghana (1996) (exchange rate and trade reforms)	1. Democratic Republic of the Congo (creating conditions for growth and economic diversification)
Sustainable growth	1. Nigeria (non-oil based sustainable growth) 2. Ghana (help create conditions for environmentally sustainable growth)	1. Nigeria (achieving sustainable and inclusive non-oil growth) 2. South Africa (regional integration) 3. Guinea (fostering sustainable and equitable growth) 4. Burundi (promoting sustainable and broad-based economic growth)
Broad based growth	1. Mali (macroeconomic stability, private sector development, agriculture and rural development) 2. Madagascar (GDP growth, agricultural productivity, natural resource management, better access to roads) 3. Malawi (labour intensive agricultural productivity and income)	1. Niger (sustainable growth that is equitably shared) 2. Senegal (promoting growth and wealth creation) 3. Mozambique (promoting equitable and broad-based growth)
Private sector economic growth	1. Mauritania (macroeconomic stability, rural and urban development) 2. Zimbabwe (expansion of private sector activity) 3. Zambia (privatizing state-owned enterprises, creating an enabling environment for private sector development) 4. Cameroon (create a climate favourable for private sector development)	1. Ghana (2012) (raise private sector competitiveness) 2. Benin (strengthening competitiveness and accelerating private sector-led growth) 3. The Gambia (enhancing productive capacity and accelerating growth and competitiveness) 4. Malawi (improve infrastructure and investment climate)

Private sector economic growth (continued)	5. Ethiopia (promotion of private sector development, privatization of state enterprises) 6. Malawi (2000) (stimulate the market-driven private sector activity, infrastructure, privatise) 7. Ghana (foster private sector development in all sectors, especially agriculture) 8. Ghana (2) (dismantle price controls) 9. Mozambique (2) (private sector development, privatisation of financial sector)	5. Republic of Congo (economic diversification and growth) 6. Ethiopia (fostering economic growth)
Labour intensive growth	1. Lesotho (agriculture, infrastructure investments, tourism)	
Poverty and Livelihoods		
Human development/ Poverty reduction	1. Mozambique 2. Uganda (health, education, water, and sanitation) 3. Mali (education, health, raise living standards) 4. Madagascar (increase pro-poor expenditures, education, health care, water and sanitation, protect vulnerable groups) 5. Malawi (2006) (education, health care, safety net, HIV/AIDS, food security) 6. Malawi (2000) (human capital development and slow population growth) 7. Senegal (living conditions, women, decentralised institutions) 8. Mauritania (education, health, infrastructure, gender) 9. Zimbabwe (equity and poverty)	1. South Africa (urban and rural development) 2. Senegal (protecting human development/ shared growth 3. Republic of Congo (poverty reduction through improved basic services) 4. Ethiopia (reducing vulnerability)

Human development/ Poverty reduction (continued)	10. Eritrea (develop the human resource base of the country) 11. Lesotho (health care, education) 12. Cameroon (alleviate poverty) 13. Ethiopia (increased access to basic education and health, and to the road network) 14. Kenya (poverty reduction) 15. Burkina Faso (investment in human capital; education, particularly for girls) 16. Ghana (2000) (accelerate action on poverty reduction) 17. Ghana (1996) (general poverty reduction) 18. Mozambique (reducing poverty through agricultural growth)	
Deliver social services, community empowerment	1. Nigeria 2. Senegal (capacity building, education, health services, infrastructure) 3. Eritrea (social fund programs for poor communities) 4. Zambia (direct assistance to poor and vulnerable) 5. Lesotho (strengthen the safety net, food security, programs to vulnerable groups) 6. Ethiopia (development of agriculture and food security) 7. Kenya (education, women's health and employment opportunities) 8. Togo (basic health, education and urban services)	1. Nigeria (sustained improvement in access and utilization of human development services) 2. South Africa (improved service delivery) 3. Ghana (improve human development outcomes) 4. Guinea (increase equitable access to social services) 5. Mauritania (develop human resources and generalize access to basic social services) 6. Democratic Republic of the Congo (health and education)

Deliver social services, community empowerment (continued)		7. Niger (human capital and access to quality social services) 8. Benin 9. Malawi (decrease vulnerability to HIV/AIDS and malnutrition) 10. Zambia (cross-cutting theme: governance and HIV measures) 11. Burundi (improving access to social services and consolidating social stability) 12. Ethiopia (improving access to and quality of basic service delivery) 13. Mozambique (ensuring equitable access to key services)
Refugees	1. Rwanda (recovery, resettling and integrating refugees, and in reconstruction)	
Governance		
Governance	1. Mozambique 2. Uganda (connected to decentralisation/anti-corruption) 3. Madagascar (anti-corruption, public sector management) 4. Kenya (reallocate public expenditures towards the basic functions of the Government, civil- service reform, public enterprise reform and privatization) 5. Ghana (program for capacity building with good governance)	1. Nigeria (governance for results) 2. Ghana (strengthen governance and civil responsibility) 3. Guinea (strengthen governance and institutional and human capacity) 4. Mauritania (good governance and capacity building) 5. Liberia (rebuilding core state functions)

Governance (continued)		6. Niger (cross cutting issues: population growth and good governance) 7. Burundi (cross-cutting pillar: strengthening governance) 8. Ethiopia (fostering improved governance) 9. Mozambique (increasing accountability and public voice)
Public finance management	1. Mali (and governance) 2. Lesotho (Civil service reform, privatise, planning and budgeting) 3. Cameroon (consolidate the benefits of the devaluation)	
Institutional development	1. Mauritania (governance, public sector reform, decentralization, strengthen legal system) 2. Lesotho (institutional capacity, poverty reduction program, decentralize administrative authority) 3. Malawi (enhance institutional capacity in the country) 4. Liberia (rebuild state functions and institutions) 5. Kenya (strengthening accountability of institutions, anti-corruption)	1. Benin (promoting better governance and strengthening institutional capacities) 2. Gambia (strengthening the institutional framework for economic management and public service delivery) 3. Zambia (institutional capacity enhancement) 4. Republic of Congo (governance and institutional capacity-building)

Public sector restructuring	1. Kenya (redefinition of the core functions of government, privatization) 2. Zambia (public sector efficiency and governance) 3. Uganda (budget reform, capacity building, civil service reform, decentralization, and governance) 4. Rwanda (institutions for economic development, key policy, legal, and institutional measures to stabilize the economy)	
Budgeting	1. Kenya (balanced budget for central government) 2. Rwanda (improve budget management to restart the growth process post-conflict)	1. Malawi (sustain improvement in fiscal discipline, budget execution & civil service accountability)
Infrastructure		
Infrastructure and development	1. Liberia 2. Eritrea (to support private sector development and the provision of public services) 3. Mozambique (2) (infrastructure development) 4. Ghana (2) (infrastructure)	1. Liberia (rehabilitating infrastructure)
Sector Reform		
Agriculture	1. Kenya (privatisation) 2. Tanzania (agriculture and growth) 3. Burkina Faso (agriculture, food production, natural resources management) 4. Mozambique (agriculture and rural development)	1. Malawi (improve smallholder productivity and integration into agro-processing)

General reform	1. Tanzania (sectoral strategies for power, transport, and finance) 2. Zambia (sector investment programs)	1. Senegal (maximizing urban/rural synergies)
Transportation	1. Togo (essential transport and energy infrastructure)	
Monitoring and Evaluation		
General		1. Mauritania (reinforce pilot programs, monitoring and evaluation and coordination)

Bibliography

Ainsworth, M. and S. Carvahlo (n.d.) *Preparing High-Quality Implementation Completion and Results Reports*, [Online: ieg.worldbankgroup.org/Data/eval_week_ preparing_icr.pptx], Washington, DC: World Bank Group.

Alesina, A. and D. Dollar (2000) Who Gives Foreign Aid to Whom and Why? *Journal of Economic Growth*, 5: 33–63.

Arksey, H. and P. Knight (1999) *Interviewing for Social Scientists: An Introductory Resource with Examples*, London: Sage Publications.

Arndt, C., S. Jones, and F. Tarp (2010) *Aid, Growth, and Development: Have We Come Full Circle?*, [Online: http://www.wider.unu.edu/publications/working-papers/2010/en_GB/wp2010-96/], Helsinki: UNU World Institute for Development Economics Research (UNU-WIDER).

Audenhove, L. (2009) Expert Interviews and Interview Techniques for Policy Analysis, [Online: http://www.ies.be/files/060313%20Interviews_VanAudenhove.pdf], Brussels: Vrije Universiteit Brussel.

Baker, J. (2000) *Evaluating the Impact of Development Projects on Poverty: A Handbook for Practitioners*, The World Bank: Washington, DC.

Bamberger, M. (2000) The Evaluation of International Development Programs: A View from the Front, *American Journal of Evaluation*, 21 (1): 95-102.

Banerjee, A. V. and E. Duflo (2009) The Experimental Approach to Development Economics, *Annual Review of Economics*, 1: 151-78.

Banerjee, A. V. and E. Duflo (2011) *Poor Economics: A Radical Rethinking of the Way to Fight Global Poverty.* New York: Public Affairs.

Banerjee, A. V. et al. (2007) *Making Aid Work.* Cambridge, MA, USA: MIT Press.

Baran, P. (1957) *The Political Economy of Growth.* New York: Penguin Books.

Barder, O. (2009) *What Is Poverty Reduction?* Working Paper 170, [Online: http://www.cgdev.org/files/1421599_file_Barder_Poverty_Reduction.pdf], Washington, DC: Center for Global Development.

Barry, A. (2006) Technological Zones, *European Journal of Social Theory*, 9 (2): 239–253.

Bathelt, H. and J. Glückler (2003) Toward a relational economic geography, *Journal of Economic Geography*, 3: 117-144.

Bauer, P. (1971) *Dissent on Development*. London: Weidenfield and Nicolson.

Bebbington, A. (2003) Global networks and local developments: agendas for development geography, *Tijdschift voor Economische en Sociale Geografie*, 94 (3): 297-309.

Biggs, S. and S. Smith (2003) A Paradox of Learning in Project Cycle Management and the Role of Organisational Culture, *World Development*, 31: 1743-1757.

Binnendijk, A.L. (1989). Donor Agency Experience with the Monitoring and Evaluation of Development Projects, *Evaluation Review* 13 (3): 206-222.

Birdsall, N. and F. Fukuyama (2011). The Post-Washington Consensus: Development After the Crisis, [Online: http://www.foreignaffairs.com/articles/67456/nancy-birdsall-and-francis-fukuyama/the-post-washington-consensus], *Foreign Affairs,* March/April.

Birdsall, N. (ed.) (2006) *Rescuing the World Bank.* Washington, DC: Center for Global Development.

Blalock, A. B. (1999) Evaluation Research and the Performance Management Movement: From Estrangement to Useful Integration? *Evaluation*, 5 (2): 117-149.

Boggs, J. S. and N. M. Rantisi (2003) The 'relational turn' in economic geography, *Journal of Economic Geography*, 3: 109-116.

Boone, P. (1996) Politics and the effectiveness of foreign aid, *European Economic Review*, 40 (2): 289-329.

Boone, P. (1994) Down the Rathole: Many argue that foreign aid is a waste of money: they may be right. *The Economist,* 10 December. (Note: *The Economist* does not give citations to authors, but Boone has been credited as the author of this article).

Boschma, R. and R. Martin (2007) Editorial: Constructing an evolutionary economic geography, *Journal of Economic Geography*, 7 (5): 537-548.

Bourguignon, F. and M. Sundberg (2007) Aid Effectiveness: Opening the Black Box, *The American Economic Review*, 97 (2): 316-321.

Bretton Woods Project (2010) World Bank Country Assistance Strategies, [Online: http://www.brettonwoodsproject.org/2010/04/art-566117/],London: Bretton Woods Project.

Brigg, M. (2002) Post-development, Foucault and the colonisation metaphor, *Third World Quarterly*, 23 (3): 421-436.

Broad, R. (2007) Knowledge Management: a case study of the World Bank's research department, *Development in Practice*, 17 (4-5): 700-708.

Broad, R. (2006) Research, Knowledge, and the Art of 'Paradigm Maintenance': The World Bank's Development Economics Vice-Presidency (DEC). *Review of International Political Economy*, 13 (3): 387-419.

Brown, A., Y. Gabriel, and S. Gherardi (2009) Storytelling and Change: An Unfolding Story, *Organization*, 16 (3): 323–333.

Brunner, R. (2004) Context-sensitive monitoring and evaluation for the World Bank, *Policy Sciences*, 37: 103–136.

Bueno de Mesquite, B. and A. Smith (2007) Foreign Aid and Policy Concessions, *Journal of Conflict Resolution*, 51 (2): 251-284.

Burawoy, M. et al. (eds) (2000) *Global Ethnography: Forces, Connections and Imaginations in a Postmodern World.* Berkeley, CA: University of California Press.

Burnside, C. and D. Dollar (2000) Aid, policies, and growth, *American Economic Review*, 90 (4): 847–868.

Campbell J., A note on growth poles, *Growth and Change*, 5 (2): 43-45.

Carvalho, S. (n.d.) Project Ratings: Connects and Disconnects, [Online: *ieg.worldbankgroup.org/Data/eval_week_project_ratings.pptx*], Washington, DC: World Bank.

Carlin, W. (2013) Economics explains our world – but economics degrees don't, [Online: http://www.ft.com/cms/s/0/74cd0b94-4de6-11e3-8fa5 00144feabdc0.html#axzz3 SZkAHvE3], Opinions, *Financial Times*, 17 November.

Cassen, R. et al. (1986) *Does aid work? Report to an intergovernmental task force.* Oxford: Oxford University Press.

Chelimsky, E. and W. R. Shadish (eds.) (1998) *Evaluation for the 21st Century: A handbook.* Thousand Oaks, California, USA: Sage Publications.

Chouinard, J. A. (2013) The Case for Participatory Evaluation in an Era of Accountability, *American Journal of Evaluation*, 34 (2): 237-253.

Christie, C. (2007) Reported Influence of Evaluation Data on Decision Makers' Actions: An Empirical Examination, *American Journal of Evaluation*, 28 (1): 8-25.

Clemens, M. et al. (2004) *Counting Chickens When They Hatch: Timing and the Effects of Aid on Growth,* [Online: http://www.cgdev.org/files/2744_file_CountingChickens FINAL3.pdf], Washington, DC: Center for Global Development.

Collier, P. (2007) *The Bottom Billion: Why the Poorest Countries are Failing and What Can Be Done About It.* New York: Oxford University Press.

Collier, P. and D. Dollar (2002) Aid allocation and poverty reduction, *European Economic Review*, 46 (8): 1475–1500.

Collier, P. and D. Dollar (2001) Can the World Cut Poverty by Half? How Policy Reform and Effective Aid can Meet International Development Goals, *World Development*, 29 (11): 1787-2002.

Conlin, S. and R. L. Stirrat (2008) Current Challenges in Development Evaluation, *Evaluation*, 14 (2): 193-208.

Corbridge, S. (1986) *Capitalist World Development: A Critique of Radical Development Geography.* New Jersey, USA: Rowman & Littlefield.

Corbridge, S., N. Thrift, and R. Martin (1994) *Money, Power and Space.* New York: Wiley.

Cormode, L. and A. Hughes (1999) Introduction: the economic geographer as a situated researcher of elites, *Geoforum*, 1999, 30 (4): 299-300.

Cowen, M. P. and R. W. Shenton (1996) *Doctrines of Development.* London and New York: Routledge.

Clark, G. L. (1998) Stylised Facts and Close Dialogue: Methodology in Economic Geography, *Annals of the Association of American Geographers*, 88 (1): 73-87.

Clark, G. L., M. S. Gertler, and M. P. Feldman (2000) *The Oxford Handbook of Economic Geography*. Oxford: Oxford University Press.

Clarke, A. and R. Dawson (1999) *Evaluation Research: An Introduction to Principles, Methods and Practice*. London: Sage Publications.

Cousins, J. B. and J. A. Chouinard (2012) *Participatory Evaluation Up Close: A Review and Integration of Research-Based Knowledge.* Connecticut, USA: Information Age Publishing.

Cracknell, B. E. (2000) *Evaluating Development Aid: Issue, Problems and Solutions.* New Delhi: Sage Publications.

Cracknell, B. E. (1996) Evaluating Development Aid Strengths and Weaknesses, *Evaluation*, 2: 23-33.

Crush, J. (1995) *Power of Development*. New York, NY: Routledge.

Czarniawska, B. (1997) *Narrating the Organization: Dramas of Institutional Identity*. Chicago: University of Chicago Press.

Czarniawska, B. (1998) A Narrative Approach to Organisation Studies, *Qualitative Research Methods,* 43: 1-87.

Dale, R. (1998) *Evaluation Frameworks for Development Programmes and Projects.* Sage Publications, New Delhi.

Davidson, E. J. (2000) Ascertaining Causality in Theory-Based Evaluation, *New Directions for Evaluation*, 87: 17-26.

Davies, R. and J. Dart (2005) *The Most Significant Change Technique: A Guide to its Use*, [Online: http://www.mande.co.uk/docs/MSCGuide.pdf], London: CARE International.

Deichmann, U., G. Gill, and C. C. Goh (2010) *World Development Report 2009:* A Practical Economic Geography, *Economic Geography*, 86 (4): 371-380.

Deininger, K., L. Squire, and S. Basu (1998) Does economic analysis improve the quality of foreign assistance? *World Bank Economic Review*, 12 (3): 385–418.

Denizer, C., D. Kaufmann, and A. Kraay (2013) *Good Countries or Good Projects? Macro and Micro Correlates of World Bank Project Performance*, [Online: http://siteresources.worldbank.org/DEC/Resources/DenizerKaufmannKraayGoodCountriesorGoodProjectsApril2013.pdf], Washington, DC: World Bank.

Dollar, D. (1998) *Assessing Aid: What Work, What Doesn't, and Why?* Washington, DC and Oxford: World Bank and Oxford University Press.

Dollar, D. and V. Levin (2006) The Increasing Selectivity of Foreign Aid, 1984–2003, *World Development*, 34 (12): 2034-2046.

Dollar, D. and J. Svensson (2000) What Explains the Success or Failure of Structural Adjustment Programmes? *The Economic Journal*, 110 (466): 894-917.

Donaldson, S. I., C. A. Christie, M. M. Mark (eds.) (2009) *What Counts as Credible Evidence in Applied Research and Evaluation Practice?* Thousand Oaks, CA: Sage Publications.

Doucouliagos, H. and M. Paldam (2009) The Aid Effectiveness Literature: The Sad Results of 40 Years of Research, *Journal of Economic Surveys*, 23 (3): 433-461.

Easterly, W. (2006) *The White Man's Burden: Why the West's Efforts to Aid the Rest Have Done So Much Ill and So Little Good.* New York: Penguin Press.

Escobar, A. (1995) *Encountering Development*. New Jersey, USA: Princeton University Press.

Essex, J. (2012) The politics of effectiveness in Canada's international development Assistance," *Canadian Journal of Development Studies (Revue canadienne d'etudes du developpement)*, 33 (3): 338-355.

Essex, J. (2010) Sustainability, food security, and development aid after the food crisis: Assessing aid strategies across donor contexts, *Sustainability*, 2 (11): 3354-3382.

Essex, J. (2008) The neoliberal geopolitics of food security: The case of Indonesia, *Human Geography*, 1 (2): 14-25.

Fairclough, N. (2013) Critical discourse analysis and critical policy studies, *Critical Policy Studies*, 7 (2): 177-197.

Ferguson, J. (1994) The Anti-Politics Machine: 'Development' and Bureaucratic Power in Lesotho, *The Ecologist*, 24 (5): 176-181.

Ferraro, V. (2008) Dependency Theory: An Introduction, pp 58-64 in Secondi, G (ed) (2008) *The Development Economics Reader*, London: Routledge.

Fine, B. (2004) Examining the Idea of Globalisation and Development Critically: What Role for Political Economy? *New Political Economy*, 9 (2): 213-231.

Fitz-Gibbon, C. T. (2002) Evaluation in an Age of Indicators: Challenges for Public Sector Management, *Evaluation*, 8 (1): 140-148.

Fletcher, G. (1989) *The Keynesian Revolution and Its Critics: Issues of Theory and Policy for the Monetary Production Economy*. New York: Palgrave MacMillan.

Flyvbjerg, B. (2011) Case Study, pp. 301-317 in Denzin, NK. and YS. Lincoln (eds.) (2011) *The Sage Handbook of Qualitative Research.* Thousand Oaks, CA: Sage Publications.

Flyvbjerg, B. (2006) Five Misunderstandings about Case-Study Research, *Qualitative Inquiry*, 12 (2): 219-245.

Foucault, M. and C. Gordon (ed.) (1980) *Power/Knowledge: Selected Interviews and Other Writings, 1972-1977*. New York: Pantheon Books.

Fujita, M. and P. Krugman (2004) The new economic geography: Past, present and the future, *Regional Science*, 83: 139–164.

Gabriel, Y. (2000) *Storytelling in Organizations: Facts, Fictions, and Fantasies.* Oxford: Oxford University Press.

Gelb, A. (2010) *How Can Donors Create Incentives for Results and Flexibility for Fragile States?* Center for Global Development Working Paper 227 [Online: http://www.cgdev.org/files/1424541_file_Gelb_IDA_flexibility_FINAL.pdf] Washington, D.C.: Center for Global Development.

Gertler, M. S. (2010) Rules of the Game: The Place of Institutions in Regional Economic Change, *Regional Studies*, 44 (1): 1–15.

Gertler, M. S. (2003) Tacit knowledge and the economic geography of context, or The undefinable tacitness of being (there), *Journal of Economic Geography*, 3 (1): 75-99.

Gilbert, C. and D. Vines (eds.) (2000) *The World Bank: Structures and Policies*. Cambridge: Cambridge University Press.

Gidwani, V. (2002) The unbearable modernity of development? Canal irrigation and development planning in Western India, *Progress in Planning*, 58 (1): 1-80.

Glassman, J. and A. I. Samatar (1997) Development geography and the third-world state, *Progress in Human Geography*, 21 (2): 164-198.

Grabher, G. (2006) Trading routes, bypasses, and risky intersections: mapping the travels of 'networks' between economic sociology and economic geography, *Progress in Human Geography*, 30 (2): 163–189.

Grasso, P., S. S. Wasty, and R. V. Weaving (eds.) (2003) *World Bank Operations Evaluation Department: The First 30 Years*, [Online: https://ieg.worldbankgroup.org/Data/reports/oed_30yrs.pdf], Washington, DC: The World Bank.

Grasso, P. (2003) What Makes an Evaluation Useful? Reflections from Experience in Large Organizations, *American Journal of Evaluation*, 24: 507-514.

Grove, N. J. and A. B. Zwi (2008) Beyond the Log Frame: A New Tool for Examining Health and Peacebuilding Initiatives, *Development in Practice*, 18 (1): 66-81.

Gulrajani, N. (2011) Transcending the great foreign aid debate: managerialism, radicalism and the search for aid effectiveness, *Third World Quarterly*, 32 (2): 199-216.

Hansen, H. and F. Tarp (2000) Policy Arena: Aid Effectiveness Disputed, *Journal of International Development*, 12: 375-398.

Harriss, J. (2002) *Depoliticizing Development: The World Bank and Social Capital.* London: Anthem Press.

Harper, R. (2000) The Social Organization of the IMF's Mission Work: An examination of international auditing, pp. 21-54 in M. Startern, ed. *Audit Cultures, Anthropological studies in accountability, ethics and the academy*, London and New York: Routledge.

Hart, G. (2010) Redrawing the Map of the World? Reflections on the *World Development Report 2009*, *Economic Geography*, 86 (4): 341-350.

Hart, G. (2001) Development critiques in the 1990s: culs de sac and promising paths *Progress in Human Geography*, 25 (4): 649-658.

Harvey, D. (1989) *The Condition of Postmodernity: An Enquiry into the Origins of Cultural Change.* Oxford: Blackwell.

Harvey, D. (2009) Reshaping Economic Geography: *The World Development Report 2009, Development and Change*, 40 (6): 1269–1277.

Heider, C. (2014) Who Evaluates the Evaluators?, *Transforming Development Through Evaluation* Blog, 7 May [Online: https://ieg.worldbankgroup.org/blog/who-evaluates-evaluators] Washington, DC: World Bank.

Heider, C. (2013) The Fight for Independence, *Transforming Development Through Evaluation* Blog, 1 September, [Online: https://ieg.worldbankgroup.org/blog/fight-independence], Washington, DC: World Bank.

Holvoet, N. and R. Renard (2007) Monitoring and evaluation under the PRSP: Solid rock or quicksand? *Evaluation and Program Planning*, 30 (1): 66–81.

House, E. (2001) Unfinished Business: Causes and Values, *American Journal of Evaluation*, 22 (3): 309-315.

Howes, S., S. Otor, and C. Rogers (2011) Does the World Bank have a micro-macro paradox, or do the data deceive? Development Policy Centre [Online: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1885227##], Canberra: Crawford School of Public Policy, The Australian National University.

Huchzermeyer, M. and A. Karam (2006) *Informal Settlements: A Perpetual Challenge?* Cape Town: Juta and Company Ltd.

Huntington, S. P. (1968) *Political Order in Changing Societies*, New Haven: Yale University Press.

Hyden, G. (2008) After the Paris declaration: taking on the issue of power, *Development Policy Review*, 26 (3): 259-274.

Isham, J., D. Kaufmann, and L. Pritchett (1997) Civil liberties, democracy, and the performance of government projects, *World Bank Economic Review*, 11(2): 219–242.

Iverson, A. (2003) *Attribution and Aid Evaluation in International Development: A Literature Review*, [Online: http://web.idrc.ca/en/ev-32055-201-1-DO_TOPIC.html], Toronto: International Development Research Centre.

Jenkins, G. P. (1997) Project analysis and the World Bank, *American Economic Review*, 87 (2): 38–42.

Jessop, B. (2001) Institutional re(turns) and the strategic–relational approach, *Environment and Planning A*, 33 (7): 1213–1235.

Jomo, K. S. and E. Reinert (2005) *The Origins of Development Economics: How schools of economic thought have addressed development*. London: Zed Books, 2005.

Jomo, K. S. and B. Fine (2006) *The New Development Economics: After the Washington Consensus.* London: Zed Books.

Keynes, J. M. (1936) *The General Theory of Employment, Interest and Money*. London: Palgrave Macmillan.

Kilby, C. (2013) The political economy of project preparation: An empirical analysis of World Bank projects, *Journal of Development Economics*, 105 (C): 211-225.

Kilby, C. (2000) Supervision and performance: The case of World Bank projects, *Journal of Development Economics*, 62 (1): 233-259.

Kindornay, S. (2011) A new agenda for development assistance? From aid to development effectiveness, North-South Institute [Online: http://www.nsi-ins.ca/wp-content/uploads/2012/10/2011-From-aid-to-development-effectiveness.pdf] Ottawa, Ontario: The North-South Institute.

Kingdom, B. and P. Reddel (2006) *Meeting water needs in Vietnam: How engaging communities can help lead to viable projects*, [Online: https://www.openknowledge.worldbank.org/handle/10986/10732], Washington, DC: World Bank.

Kingston, C. and G. Caballero (2009) Comparing theories of institutional change, *Journal of Institutional Economics*, 5 (2): 151-180.

Knack, S. and A. Rahman (2007) Donor fragmentation and bureaucratic quality in aid recipients, *Journal of Development Economics*, 83 (1): 176-197.

Konadu-Agyemang, K. (2000) The Best of Times and Worst of Times: Structural Adjustment Programs and Uneven Development in Africa: The Case of Ghana, *The Professional Geographer*, 52 (3): 469-483.

Kosack, S. (2003) Effective aid: How democracy allows development aid to improve the quality of life, *World Development*, 31 (1): 1-22.

Kothari, U. and M. Minogue, eds. (2002) *Development Theory and Practice: Critical Perspectives*. Basingstoke, Hampshire: Palgrave, St. Martin's Press.

Krippendorff, K. (2004) *Content Analysis: An Introduction to Its Methodology.* Thousand Oaks, CA: Sage Publications.

Krugman, P. (1998) What's new about the new economic geography? *Oxford Review of Economic Policy,* 14 (2): 7-17.

Kvale, S. (1996) *InterViews: An Introduction to Qualitative Research Interviewing*. London: Sage Publications.

Kvale, S. (1983) The qualitative research interview: A phenomenological and a hermeneutical mode of understanding, *Journal of Phenomenological Psychology*, 14 (2): 171-196.

Lancaster, C. (2007) *Foreign Aid: Diplomacy, Development, and Domestic Politics*. Chicago: The University of Chicago Press.

Larrain, J. (1989) *Theories of Development: Capitalism, Colonialism and Dependency*. London: Polity Press.

Latour, B. (1979) *Laboratory Life: The Social Construction of Scientific Facts*. London: Sage Publications.

Lawson, V. (2010) Reshaping Economic Geography? Producing Spaces of Inclusive Development, *Economic Geography*, 86 (4): 351-360.

Lawson, V. (ed) (2007) *Making Development Geography: Human Geography in the Making.* New York: Hodder Education.

Leeuw, F. L. and J. E. Furubo (2008) Evaluation Systems: What are they and why study them? *Evaluation*, 14 (2): 157-169.

Lehmann, D. (1997) An opportunity lost: Escobar's deconstruction of development, *Journal of Development Studies*, 33 (4): 568-578.

Li, T. M. (2007) *The Will to Improve: Governmentality, Development and the Practice of Politics*, North Carolina: Duke University Press.

Linn, J. (2012) *Evaluating the Evaluators: Some Lessons from a Recent World Bank Self-Evaluation* [Online: http://www.brookings.edu/research/opinions/2012/02/21-world-bank-evaluation-linn], Scaling Up Development Series, 21 September, Washington, DC: Brookings Institution.

Lipset, S. M. (1959) Some Social Requisites of Democracy: Economic Development and Political Legitimacy, *American Political Science Review,* 53 (1): 69-105.

Liverani, A. and H. Lundgren (2007) Evaluation Systems in Development Aid Agencies: An Analysis of DAC Peer Reviews 1996 –2004, *Evaluation,* 13 (2): 241-256.

Locke, T. (2004) *Critical Discourse Analysis.* London: Continuum International Publishing Group.

Lopez-Acevedo, G., P. Krause, and K. Mackay, eds. (2012) *Building Better Policies: The Nuts and Bolts of Monitoring and Evaluation Systems.* Washington, DC: The World Bank.

Lyne, M. M., D. L. Nielson, and M. J. Tierney (2006). Who delegates? Alternative models of principals in development aid, pp. 41-76 in Hawkins et al. (eds.), *Delegation and Agency in International Organizations*. Cambridge: Cambridge University Press.

MacKinnon, D. et al. (2009) Evolution in Economic Geography: Institutions, Political Economy, and Adaptation, *Economic Geography,* 85: 129-150.

Martin, R. (2008) Institutional Approaches in Economic Geography, pp. 77-94 in Sheppard, E and T Barnes, eds., *A Companion to Economic Geography.* Oxford: Blackwell Publishing.

Martin, L. L. and B. A. Simmons (1998) Theories and empirical studies of international institutions, *International Organization,* 52 (4): 729-757.

Martin, R. and P. Sunley (2011) The new economic geography and policy relevance, *Journal of Economic Geography,* 11: 357–369.

Massey, D. (1994) A global sense of place, pp 146-156 in *Space, place and gender,* Minneapolis: University of Minnesota Press.

Mawdsley, E. (2007) The Millennium Challenge Account: Neo-liberalism, poverty and security, *Review of International Political Economy*, 14 (3): 487-509.

Maxwell, S. (1997) The Use of Matrix Scoring to Identify Systemic Issues in Country-Programme Evaluation, *Development in Practice,* 7 (4): 408-415.

McDowell, L. (1998) Elites in the City of London: some methodological considerations, *Environment and Planning A,* 30 (12): 2133-2146.

McDowell, L. (1992) Doing Gender: Feminism, Feminists and Research Methods in Human Geography, *Transactions of the Institute of British Geographers,* 17 (4): 399-416.

McGregor, A. (2009) New Possibilities? Shifts in Post-Development Theory and Practice, *Geography Compass*, 3 (5): 1688-1702.

McLafferty, S. (1995) Counting for women, *The Professional Geographer*, 47 (4): 436-442.

Mekasha, T.J. and F. Tarp (2011) *Aid and growth: What meta-analysis reveals,* Working Paper No. 2011/22. Helsinki: UNU-Wider.

Merry, S. E. (2011) Measuring the World: Indicators, Human Rights, and Global Governance, *Current Anthropology*, 52 (S3): S83-S95.

Mikecz, R. (2012) Interviewing Elites: Addressing Methodological Issues, *Qualitative Inquiry*, 18 (6): 482-493.

Miles, M. and J. Crush (1993) Personal Narratives as Interactive Texts: Collecting and Interpreting Migrant Life-Histories, *Professional Geographer*, 45 (1): 84-94.

Milner, H. and D. Tingley (2010) The Political Economy of US Foreign Aid: American Legislators and the Domestic Politics of Aid, *Economics and Politics*, 22 (2): 200-232.

Mkandawire, T. and C. Soludo (1999) *Our Continent, Our Future: African Perspectives on Structural Adjustment*. Dakar: Council for the Development of Social Science Research in Africa.

Morgan, A. (2010) Discourse analysis: An overview for the Neophyte Researcher, *Journal of Health and Social Care Improvement*, 5: 1-7.

Mosely, P., J. Hudson, and S. Horrell (1987) Aid, the Public Sector and the Market in Less Developed Countries, *The Economic Journal*, 97 (387): 616-641.

Mosley, P. (1986) Aid Effectiveness: the micro-macro paradox, *Institute of Development Studies Bulletin*, 17 (2): 22-27.

Mosse, D. (2005) *Cultivating Development: An Ethnography of Aid Policy and Practice.* London: Pluto Press.

Moyo, D. (2009) *Dead Aid: Why Aid Is Not Working and How There Is a Better Way for Africa.* New York: Farrar, Straus, and Giroux.

Mullings, B. (1999) Insider or outsider, both or neither: Some dilemmas of interviewing in a cross-cultural setting, *Geoforum*, 30 (4): 337-350.

Myrdal, G. (1957) *Economic Theory and Underdeveloped Regions*. London: Duckworth.

Neuendorf, K. (2002) *The Content Analysis Guidebook*. London: Sage Publications.

Pawson, R. (2003) Nothing as practical as a good theory, *Evaluation*, 9 (4): 471-490.

Peck, J. and E. Sheppard (2010) Worlds Apart? Engaging with the World Development Report 2009, *Economic Geography*, 86 (4): 331-340.

Peet, R. and E. Hartwick (2009) *Theories of Development: Contentions, Arguments, Alternatives*. New York, NY: Guilford Press.

Perroux, F. (1955) A Note on the Notion of the Growth Pole. *Economic Appliqué*, 1 and 2: 307-320.

Picciotto, R. (2009) Development Effectiveness: An Evaluation Perspective, pp 180-211 in G Mavrotas and M McGillivray (eds.) *Development Aid: A Fresh Look.* Tokyo: United Nations University.

Picciotto, R. (2011) The logic of evaluation professionalism, *Evaluation*, 17 (2): 165-180.

Pieterse, J. N. (2010) *Development Theory.* London: Sage Publications.

Pieterse, J. N. (2000). After post-development, *Third World Quarterly*, 21 (2): 175–191.

Pigg, SL. (1992) Inventing social categories through place: Social representations and development in Nepal, *Comparative Studies in History and Society*, 34 (3): 491-513.

Powell, W. W. and P. J. DiMaggio, eds. (1991) *The New Institutionalism in Organizational Analysis.* Chicago: University of Chicago Press.

Power, M. (1999) *The audit society: rituals of verification.* Oxford: Oxford University Press.

Power, M. (2003) Auditing and the production of legitimacy, *Accounting, Organizations and Society*, 28 (4): 379-394.

Quarles van Ufford, P. and A. Kumar Giri, eds. (2003) *A Moral Critique of Development: In search of global responsibilities.* London: Routledge.

Radelet, S. (2006) *A primer on foreign aid.* Washington, DC: Center for Global Development.

Rahmena, M. and V. Bawtree (1997) *The Post-Development Reader*. London: Zed Books.

Ravallion, M. (2008) *Evaluation in the Practice of Development.* Development Research Group Working Paper, Washington, DC: World Bank.

Rebien, C. (1996) *Evaluating Development Assistance in Theory and in Practice*. Aldershot: Avebury, Ashgate Publishing.

Ricardo, D. (1817) *The Principles of Political Economy and Taxation.* London: John Murray.

Richards, D. (1996) Elite Interviewing: Approaches and Pitfalls, *Politics*, 16 (3): 199-204.

Rist, G. (2004) *The History of Development: From Western Origins to Global Faith*. London: Zed Books.

Rist, G. (2007) Development as a buzzword, *Development in Practice*, 17 (4): 485-491.

Rodríguez-Pose, A. (2010) Economic geographers and the limelight: The reaction to the 2009 World Development Report, *Economic Geography*, 86 (4): 361-370.

Rodrik, D. (2006) Goodbye Washington Consensus, Hello Washington Confusion? A Review of the World Bank's Economic Growth in the 1990s: Learning from a Decade of Reform, *Journal of Economic Literature*, 44 (4): 973-987.

Roland-Holst, D. and F. Tarp (2004) New Perspectives on Aid Effectiveness, pp. 43-63 in Stern, N., B. Tungodden, and I. Kolstad (eds.) *Toward Pro-Poor Policies: Aid, Institutions and Globalization*, Washington, DC: World Bank.

Roodman, D. (2007) *Macro Aid Effectiveness Research: A Guide for the Perplexed*, Working Paper Number 134, Washington DC: Center for Global Development.

Rose, G. (1997) Situating Knowledges: positionality, reflexivities and other tactics, *Progress in Human* Geography, 21 (3): 305-320.

Rossi, P., H. Freeman, and M. Lipsey (1999) *Evaluation: A Systemic Approach.* Thousand Oaks, CA: Sage Publications.

Rostow, W. W. (1960) *The Stages of Economic Growth: An Anti-Communist Manifesto.* Cambridge: Cambridge University Press, 1960.

Sabot, E. (1999) Dr Jekyl, Mr H(i)de: the contrasting face of elites at interview, *Geoforum,* 30 (4): 329-335.

Sachs, W. (ed.) (1992) *The development dictionary: a guide to knowledge as power.* London: Zed Books.

Saldaña, J. (2010) *The Coding Manual for Qualitative Researchers*. London: Sage Publications.

Schaumburg-Müller, H. (2007) Use of Aid Evaluation from an Organizational Perspective, *Evaluation,* 11 (2): 207-222.

Schneiberg, M. and E. S. Clemens (2006) The Typical Tools for the Job: Research Strategies in Institutional Analysis, *Sociological Theory*, 24 (3): 195-227.

Schoenberger, E. (1992) Self-criticism and self-awareness in research: a reply to Linda McDowell, *The Professional Geographer*, 44 (2): 215-218.

Schwandt, T. A. (2009) Toward a Practical Theory of Evidence in Evaluation, pp. 197-213 in Donaldson, S. I., C. A. Christie, M. M. Mark (2009) *What Counts as Credible Evidence in Applied Research and Evaluation Practice?* Thousand Oaks, CA: Sage Publications.

Scott, W. (2001) *Institutions and Organizations: Ideas and Interests, second edition.* Thousand Oaks, CA: Sage Publications.

Secondi, G (ed) (2008) *The Development Economics Reader*, Abingdon and New York: Routledge.

Sen, A. (1999) *Development as Freedom*. New York: Alfred A. Knopf.

Simon, D. (2007) Beyond antidevelopment: discourses, convergences, practices, *Singapore Journal of Tropical Geography,* 28 (2): 205-218.

Singer, H. W. (1949) Economic Progress in Underdeveloped Countries, *Social Research,* 16 (1): 1-11.

Smith, K. E. (2006) Problematising power relations in 'elite' interviews, *Geoforum,* 37 (4): 643-653.

Smith, A. (1776) *The Wealth of Nations.* London: W. Strahan and T. Cadell.

Soederberg, S. (2004) American empire and 'excluded states: the Millennium Challenge Account and the Shift to pre-emptive development, *Third World Quarterly,* 25 (2): 279-302.

Solow, R. (1956) A Contribution to the Theory of Economic Growth, *The Quarterly Journal of Economics*, 70 (1): 65-94.

Stiglitz, J. E., A. Sen, and J. P. Fitoussi, J. (2009) *Report by the Commission on the Measurement of Economic Performance and Social Progress,* [Online: www.stiglitz-sen-fitoussi.fr/documents/rapport_anglais.pdf], Commission on the Measurement of Economic Performance and Social Progress.

Storper, M. (2011) Why do regions develop and change? The challenge for geography and economics, *Journal of Economic Geography*, 11 (2): 333-346.

Strathern, M. ed. (2000) *Audit Cultures, Anthropological studies in accountability, ethics and the academy*. London and New York: Routledge.

Subramanian A. and R. Ragan (2008) Aid and Growth: What Does the Cross-Country Evidence Really Show? *The Review of Economics and Statistics*, 90 (4): 643-665.

Thomas, G. (2011) A Typology for the Case Study in Social Science Following a Review of Definition, Discourse, and Structure, *Qualitative Inquiry*, 17 (6): 511-521.

Tierney, M. et al. (2011) More Dollars than Sense Refining Our Knowledge of Development Finance Using AidData, *World Development*, 39 (11): 1891-1906.

Todaro, M. and S. Smith (2006) *Economic Development.* Boston: Pearson Addison Wesley.

Uitto, J. I. (2014) Evaluating environment and development: Lessons from international cooperation, *Evaluation*, 20 (1): 44-57.

Van Grunsven, L. and A. van Westen (2003) Globalisation and Locality in Development Geography: Introduction to the Dossier, *Tijdschift voor Economische en Sociale Geografie*, 94 (3): 291-296.

Wane, W. (2004) The Quality of Foreign Aid: Country Selectivity or Donors' Incentives? World Bank Policy Research Working Paper 3325, *World Bank Policy Research*, Washington, DC: World Bank.

Weiss, C. (1997) Theory-based evaluation: Past, present, and future, *New Directions for Evaluation Special Issue: Progress and Future Directions in Evaluation: Perspectives on Theory, Practice, and Methods*, 76: 41-55.

Williamson, J. (1990) What Washington Means by Policy Reform, chapter 2 in Williamson, ed. *Latin American Adjustment: How Much Has Happened?* Washington, DC: Pearson Institute for International Economics.

Williamson, J. (2000) What Should the World Bank Think about the Washington Consensus? *The World Bank Research Observer*, 15 (2): 251-264.

Wilson, J. (2011) Colonising Space: The New Economic Geography in Theory and Practice, *New Political Economy*, 16 (3): 373-397.

Winters, M. and J. Wright (2010) The politics of effective foreign aid, *Annual Reviews of Political Science*, 13: 61–80.

Wood, B. et al. (2011) *The evaluation of the Paris declaration, final report.* Copenhagen: Danish Institute for International Studies.

Woods, N. (2006) *The Globalizers: The IMF, the World Bank, and their Borrowers.* Ithaca, NY: Cornell University Press.

World Bank (2014) About, Organization Chart, [Online: http://www.worldbank.org/content/dam/Worldbank/About/OrgChart/Bank14_07.pdf], Washington, DC: World Bank.

World Bank (2013) *World Bank Group Impact Evaluations: Relevance and Effectiveness.* Washington, D.C.: World Bank.

World Bank (2011a) *Guidance to Staff on the Country Assistance Strategy Completion Report*, [Online: http://siteresources.worldbank.org/PROJECTS/Resources/40940-1244163232994/6180403-1337885797985/CASCRGuidelinesMay2012.pdf], Washington, DC: World Bank.

World Bank (2011b) *Access to Information Policy*, [Online: https://ieg.worldbankgroup.org/Data/A2I.pdf], Washington, DC: World Bank.

World Bank (2011c) *Growth and productivity in agriculture and agribusiness : evaluative lessons from World Bank Group experience*, [Online: http://siteresources.worldbank.org/EXTGPAA/Resources/Agribusiness_eval.pdf], Washington, DC: World Bank.

World Bank (2009) *World Development Report 2009: Reshaping economic geography.* Washington DC: World Bank.

World Bank (2005) *Economic Growth in the 1990s: Learning from a Decade of Reform.* Washington, DC: World Bank, 2005.

World Bank (2000) *Good Practices in Country-led Monitoring and Evaluation Systems.* Washington, DC: World Bank.

World Bank (n.d.) *Appendix B: Guide to IEG's Country Program Evaluation Methodology*, [Online: https://ieg.worldbankgroup.org/Data/reports/chapters/brazil_cpe_appb.pdf], Washington, DC: World Bank.

World Bank (n.d.), *Evaluation Categories*, [Online: https://ieg.worldbankgroup.org/webpage/evaluations], Washington, DC: World Bank.

World Bank (n.d.), *Methodology*, [Online: http://ieg.worldbankgroup.org/methodology], Washington, DC: World Bank.

World Bank (n.d.), *Ratings*, [Online: http://ieg.worldbankgroup.org/ratings], Washington, DC: World Bank.

World Bank Independent Evaluation Group (IEG) (2013) *Independent Evaluation Group Annual Report 2013.* Washington, DC: World Bank.

World Bank Independent Evaluation Group (IEG) (2012) *Independent Evaluation Group Annual Report 2012, The Results and Performance of the World Bank Group (RAP).* Washington, DC: World Bank.

World Bank Independent Evaluation Group (IEG) (2011) *Self-Evaluation of the Independent Evaluation Group*, [Online: http://ieg.worldbank.org/Data/ieg_self_eval.pdf], Washington, D.C.: World Bank.

World Bank Operations Evaluation Department (OED) (2005a) *Harmonized Criteria for ICR and OED Evaluations*, [Online: http://ieg.worldbankgroup.org/Data/HarmonizeEvalCriteria.pdf], Washington DC: World Bank.

World Bank Operations Evaluation Department (OED) (2005b) *Country Assistance Evaluation Retrospective: An OED Self-Evaluation.* Washington, DC: World Bank.

Yannias, A. C. (2013a) IEG Evaluation Week, Personal Notes, March. Washington, DC.

Yannias, A. C. (2013b) Email to interviewees. Personal Correspondence, January. Washington, DC.

END OF DOCUMENT