Program Evaluation in the Field of International Development

Bridging Perspectives

Authors: Ernesto Campos
         Kirsten Williams

Supervisor: Natalia Semenova
Abstract

Program evaluation has carved out its value in the International Development (ID) sector as the way that International Non-Governmental Organization (INGOs) report program performance, and more importantly learn on how to improve for future programs. Its rise to prominence has been an evolution between the traditional, outcomes and participatory evaluation paradigms.

Although program evaluation is regarded as a function of management, Project Management (PM) as a discipline has been scarcely part of ID’s growth and virtually non-existent when it comes to program evaluation. In fact, although ID is a project-based sector, it largely developed its own PM practices, independent from the PM discipline. The developments of PM in ID can also be viewed in paradigm shifts from the conventional, people-centered right down to critical approaches. The only exception to this has been the creation of PM4NGOs in 2011 – a PM association geared to tailor PM knowledge for national and international Non-Governmental Organizations (NGOs).

In this backdrop, the research employs a critical literature review and data collection process to investigate if PM has contributed to program evaluation and whether program evaluation practices seen in the sector today could inform PM4NGO’s guide: PMD Pro.

The study’s findings yield that PM has fallen short in contributing to program evaluation, despite the creation of PM4NGOs. Rather, what is evidenced in the PMD Pro guide is an adherence to traditional paradigms and an inability to capture where the sector is today: in an outcomes evaluation paradigm within a people-centered PM in ID era. This study ultimately provides a series of recommendations to update PM4NGO’s contribution to program evaluation and synchronize its relevance with program evaluation practices in the sector today.

Keywords: Program Evaluation, International Development, Project Management, International Non-Governmental Organizations
Acknowledgements

We would like to express our sincere gratitude to all people who supported and contributed to this study. Especially to the M&E officers from the organizations interviewed who gave us their valuable time and opened their offices to us. We would also like to thank Professor Rugggero Golini for taking the time to talk with us and share his insights.

To our supervisor Natalia Semenova, for her keen eye and work ethic that helped us push ourselves and make this a worthwhile endeavour.

To our families, for believing in us, and being there when we needed it.

And lastly, to all twenty-nine of our MSPME friends. Thank you for the memories, the laughs and the challenges. Sharing this journey with all of you was amazing.

Ernesto Campos           Kirsten Williams

January 2016
Table of Contents

Chapter 1. Introduction ........................................................................................................... 1
  1.1 Background .................................................................................................................... 1
  1.2 Research Questions and Object of Study ................................................................. 4
    1.2.1 Research Questions .............................................................................................. 4
    1.2.2 Object of Study .................................................................................................... 5

Chapter 2. Research Methodology ....................................................................................... 7
  2.1 Theoretical Principles of Data Collection ................................................................. 7
  2.2 Ontological and Epistemological Orientations of Research .................................... 7
  2.3 Theoretical Contribution ........................................................................................... 8
  2.4 Research Approach Method ....................................................................................... 9
  2.5 Data Collection ........................................................................................................... 10
    2.5.1 Critical Literature Review .................................................................................... 10
    2.5.2 Semi-structured Interviews ................................................................................ 11
      2.5.2.1 Limitations of Interviews ............................................................................. 13
    2.5.3 Content Analysis ................................................................................................. 14
      2.5.3.1 Limitations of Content Analysis .................................................................. 15
    2.5.4 Triangulation ...................................................................................................... 16
    2.5.5 Ethical Considerations ......................................................................................... 16

Chapter 3. Theoretical Framework ...................................................................................... 18
  3.1 Concepts Definition ..................................................................................................... 18
  3.2 Assumptions and Limitations ...................................................................................... 20
  3.3 International Development ......................................................................................... 21
  3.4 PM in the field of ID .................................................................................................... 22
  3.5 Growth and Evolution of PM in ID ............................................................................ 25
  3.6 The Rise of Program Evaluation ................................................................................ 27
  3.7 Program Evaluation and PM: Bridging the Gap ....................................................... 28
  3.8 PMD Pro: A Critical Analysis of PM4NGOs Contribution to Program Evaluation ....... 32
    3.9 Bridging the Literature ............................................................................................. 34

Chapter 4. Data Presentation ............................................................................................... 36
  4.1 Semi-Structured Interviews ....................................................................................... 36
    4.1.1 INGO Interview Findings ................................................................................... 36
    4.1.2 Key informant Interview Findings ....................................................................... 42
  4.2 Content Analysis ......................................................................................................... 45
    4.2.1 Content Analysis Findings .................................................................................. 45

Chapter 5. Data Analysis ..................................................................................................... 49
  5.1 Integration of PM Practices in Program Evaluation .................................................. 49
    5.1.1 Creation of PM Tools in ID: Parallel Growth ..................................................... 49
    5.1.2 Trying to Bridge the Gap: PMD Pro .................................................................... 50
  5.3 Lessons Learned: Program Evaluation Practice ....................................................... 53
    5.3.1 Variation in Program Evaluation ......................................................................... 53

iv
5.3.2 Donor Influences on Program Evaluation .......................... 54
5.3.3 Commission External Evaluations .................................. 56
5.3.4 Adopting DAC Criteria .................................................. 57
5.3.5 Program Evaluation: In Between Paradigms ...................... 57

Chapter 6. Recommendations and Conclusions .......................... 60
  6.1 Recommendations ....................................................... 60
  6.2 Truth Criteria ............................................................. 62
    6.2.1 Validity ............................................................... 62
    6.2.2 Reliability .......................................................... 63
    6.2.3 Generalizability .................................................... 63
  6.3 Theoretical Contribution and Future Research ...................... 64
  6.4 Conclusion ................................................................... 65

References ............................................................................. 66
Interviews ................................................................................ 71
Appendixes ............................................................................. 73
  Appendix 1 .......................................................................... 73
  Appendix 2 .......................................................................... 74
  Appendix 3 .......................................................................... 75
  Appendix 4 .......................................................................... 77
  Appendix 5 .......................................................................... 78
  Appendix 6 .......................................................................... 79
  Appendix 7 .......................................................................... 80

List of Tables:
  Table 1. Organizations / people interviewed and length of each interview… 13
  Table 2. Key Characteristics of Paradigm Shifts Within Project Management in the Development Sector .................................................. 26
  Table 3. Key Characteristics of Paradigm Shifts Within Development Evaluation ................................................................. 73

List of Figures:
  Figure 1. Paradigms Within Program Evaluation and Project Management in the Development Sector .................................................. 20
  Figure 2. Key Characteristics of Paradigm Shifts Within Project Management in the Development Sector .................................................. 35
  Figure 3. Frequency of Terms within Interviews ........................................ 42
  Figure 4. Use of DAC Criteria ....................................................... 46
  Figure 5. Frequency of Specific Criteria ............................................. 46
  Figure 6. Ranking of Criteria in ToRs .................................................. 47
  Figure 7. Approaches to Evaluation ................................................... 48
  Figure 8. PMD Pro Project Phase Model ............................................. 51
  Figure 9. PMD Pro Project Phase Model ............................................. 53
# Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ID</td>
<td>International Development</td>
</tr>
<tr>
<td>PM</td>
<td>Project Management</td>
</tr>
<tr>
<td>INGOs</td>
<td>International Non-Government Organizations</td>
</tr>
<tr>
<td>BoK</td>
<td>Body of Knowledge</td>
</tr>
<tr>
<td>PM4NGOs</td>
<td>International Association for Project Management in Non-Government Organizations</td>
</tr>
<tr>
<td>PMD Pro Guide</td>
<td>Guide to the PM4NGOs BoK</td>
</tr>
<tr>
<td>PCM</td>
<td>Project Cycle Management</td>
</tr>
<tr>
<td>LFA</td>
<td>Logical Framework Approach</td>
</tr>
<tr>
<td>RBM</td>
<td>Results Based Management</td>
</tr>
<tr>
<td>ToR</td>
<td>Terms of Reference</td>
</tr>
<tr>
<td>ToC</td>
<td>Theory of Change</td>
</tr>
<tr>
<td>DAC</td>
<td>Development Assistance Committee</td>
</tr>
<tr>
<td>OECD</td>
<td>Organization for Economic Co-operation and Development</td>
</tr>
</tbody>
</table>
Chapter 1. Introduction

1.1 Background

Program evaluation is a booming practice in the field of International Development (ID) (Lennie, et al., 2015, p.326). It’s rise to prominence comes at a time when more funds than ever before are being channeled towards development assistance in ‘developing countries’, the demand for accountability is at its highest. As such, the need for more sophisticated Project Management (PM) practices to facilitate successful ID programs is paramount (Crawford & Bryce, 2002, p.363). In this context, program evaluation is regarded as a critical management function to facilitate an understanding of how programs have performed and how they can be improved upon (Fraser & Rogers, 2014, p.2). Its role is thus vital in the ID field where international non-governmental organizations (INGOs) implement programs that are too often ill fated and have thus become subject to scrutiny (Ika & Hodgson, 2014, p.1191). Even though program evaluation can be understood as an integral part of PM (Marshall and Suárez, 2013, p.1045), this research problematizes the seemingly outdated contribution to program evaluation from PM4NGOs – a PM association geared to tailor PM knowledge for national and international Non-Governmental Organizations (NGOs) – as compared to contributions from the ID field. As such, this thesis seeks to explore the extent to which PM has contributed to program evaluation and whether current program evaluation practices can inform PM4NGOs.

The management of International Development projects has become an increasing dilemma. The more the industry of helping ‘developing countries’ out of poverty, via ‘development programs’, grows (reported at $136 billion USD committed to official development assistance in 2010); the more there is a need to understand the impact of such programs (Hermano, et al., 2012, p. 22; Golini, et al., 2014, p.650). The phenomenon of ‘development management’, i.e the management of development projects and programs, is palpable. The execution of these programs implies a host of global players. It often involves international development agencies (funders of programs) and international non-governmental organizations (INGOs) (executors of programs), among other stakeholders involved in the financing, supervision, and evaluation of these programs.

What’s more, the increasing attention of management has not only spurred new debates within the scholarly community of international development, but also within the management and business administration circles. In the time period spanning from 1996 to 2006, higher education graduate programs concerning management had an overall increase of 25% in courses being offered that dealt with non-profit management education and management in development (Mirabella, 2007, p. 14S).
Central to these management conversations is the booming practice of program evaluation. (Lennie, et al., 2015, p.326). It’s rise to prominence comes at a time when more funds than ever before are being channeled towards development assistance in ‘developing countries’, the demand for accountability is at its highest and the need for more sophisticated Project Management (PM) practices to facilitate successful ID programs is paramount (Crawford & Bryce, 2002, p.363). In this context, program evaluation is regarded as a critical management function to facilitate an understanding of how programs have performed and how they can be improved upon (Fraser & Rogers, 2014, p.2). Its role is thus vital in the ID field where INGOs implement programs that are too often ill fated and have thus become subject to scrutiny (Ika & Hodgson, 2014, p.1191).

In effect, the question of how programs in the development industry are being handled has become an important area of research. It’s relevance for the project management community has also become increasingly evident. First and foremost is the fundamental fact that assistance to developing countries is provided pre-dominantly via projects and programs (Corti & Landoni, 2011, p.45). As such, numerous stakeholders from the international community are interested in understanding how funds are being used to drive impact (Crawford & Bryce, 2002, p. 364). Indeed, recent years have seen a growth in the trend of looking at how money is used in ID programs (Golini, et al., 2014, p. 650), and that is only the tip of the iceberg when it comes to evaluation and program management.

Most of the articles written up until this point about ID programs have focused on their management from a traditional PM perspective (Golini, et al., 2014; Corti & Landoni, 2011; Crawford & Bryce, 2002). That is to say, they adopt a perspective that tools and methodologies from PM can be transversally applied across different sectors. These studies advocate that “the proper use of specific methodologies and tools is critical to manage projects successfully” (Golini, et al., 2014, p.652; Ika, et al., 2010, p. 64). Professionals and academics alike have pegged this as a “one-size-fits-all” solution to ID programs (Golini, et al. 2014, p.651). Moreover they share a view that project success is exclusively a function of efficiency and to a lesser degree effectiveness.

Yet these types of tools and methodologies carry an inherent limitation: they confine the measurement of success to the iron triangle of success: time, cost and quality. In doing so, they neglect that ID programs are different in nature. Programs in the ID sector involve larger number of stakeholders that have different perspectives and complex relationships with one another. Moreover, in principal ID programs are permeated by social and “not-for-profit” modes of thinking that distinguish it from the private sector (Corti & Landoni, 2011, p.45). Previous studies by Crawford and Bryce (2003), Golini et al. (2014), Ika, et al. and Youker (2003) have recognized these types of differences and have called for a more adaptive approach when applying PM techniques, tools and methodologies to other sectors.
Among the few efforts to connect PM to ID programs in a meaningful way, the creations of PM4NGOs in 2011 has been the most concrete attempt to bridge these differences and try to be more adaptive of the tools and techniques utilized in ID programs and projects. Branded as a PM association tailoring PM knowledge for national and international Non-Governmental Organizations (NGOs), PM4NGOs has more tangibly catered to the specific characteristics of ID through the development of the PMD Pro guide. This guide acts as a more accessible alternative to the traditional standards of PM tools, by more clearly articulating how PM tools can be adapted for use in the development sector (Hermano, et al., 2013, p. 23). The guide is founded on two premises: first that project managers within the ID sector face similar challenges to those of their counterparts in other sectors. Second, that there is ample room for improving program success by abstracting lessons learned from how PM has been applied in other sectors. Currently only a study carried out by Hermano, et al. from 2013 has considered the guide and how it relates to current tools used in the management of ID programs. Yet, to date, no studies have been conducted to assess how PM has contributed to program evaluation as a growing function of ID program management. Nor has their been any research to scrutinize whether PMD Pro’s chapter on evaluation is well-informed and up to date.

With little explicit mention of ‘evaluation’ within the PM in ID literature, clues on how PM views the subject can be drawn more aptly from narratives around project success. Understanding what drives success is especially pertinent in a sector that has been recognized to experience project failure rates of up to 64% (Ika & Hodgson, 2014, p. 1184). This high rate of failure is a reminder of why the conversation around program evaluation is so crucial. What most research reveals is that ID program success is often discussed in terms of efficiency and to a lesser degree effectiveness of programs. Yet, the literature around program evaluation from the development community is far more ample. As such this study is interested in taking a closer look at the novel ideas being generated from the development community and sharing it with PM.

Among the few authors to have critically thought about evaluation from a management perspective, Crawford and Bryce (2003) provide insight into trends and dilemmas in ID program evaluation. They make an important distinction between evaluation and monitoring, the latter being far more akin to how PM talks about evaluation – from an efficiency angle (p. 366). The authors juxtapose monitoring with the notion of evaluation in the development field: where effectiveness is more important. This means going beyond the outputs of a project in terms of time, cost and quality. Rather, greater emphasis is placed on measuring the appropriateness of the program – ultimately defined by the ecological, social and economic sustainability of the initiative (Crawford & Bryce, 2003, p. 366).

Even though program evaluation can be understood as an integral part of PM (Marshall and Suárez, 2013, p.1045), this research problematizes the seemingly outdated contribution to program evaluation from PM4NGOs as compared to contributions from...
the ID field. As such, this thesis seeks to explore the extent to which PM has contributed to program evaluation and whether current program evaluation practices can inform PM4NGOs.

1.2 Research Questions and Object of Study

1.2.1 Research Questions

Before delving full immersion into the research and analysis, it is useful to clearly locate the subject of study. To do this, Whetten (1989) suggests that researchers first identify the correct constructs and justify their selection against the criterion of comprehensiveness and parsimony (p.490). In this case, constructs are limited to program evaluation, PM’s contribution to program evaluation, and ID’s contribution to program evaluation. At this point it is worth clarifying a few references. From here on out, unless otherwise stated, program evaluation will be understood as program evaluation within the ID sector and as practiced by INGOs as the main executors of programs (Marshall & Suárez, 2014, p.1037). Likewise, ‘contribution’ is conceived here as whether a body of knowledge or set of tools – coming from either traditional PM or ID – has been offered and adopted by the intended target.

The research will first and foremost consider what PM has contributed to program evaluation. It is important to consider at this point, that previous research has been scant and scattered when talking about program evaluation. Most studies link the process of evaluation with the idea of monitoring, as a tool to mitigate poor project performance (Crawford & Bryce, 2002, p.363). Much of the literature and the research carried out in terms of program evaluation have focused on how to apply existing project management tools and methodologies to ID programs and projects. Even so, this has not helped the success rates of these projects (Ika & Hodgson, 2014, p. 1184). Thus there is a need to understand what has been the actual contribution of PM to ID.

Secondly, this thesis wishes to observe what ID has contributed to program evaluation. Crawford and Bryce make a clear affirmation that currently there is a “growing volume of literature to support the demand for knowledge about monitoring and evaluation” (2002, p. 364). This claim stands on the basis that there is a lack of flexibility in the frameworks and tools existing, to accommodate for the characteristics encountered in ID programs. The underlying assumption projected by the literature is that PM’s contribution to ID has been scarce when it comes to program evaluation; largely holding on to an outdated evaluation appraisal of programs based on scope, cost and time (Hermano et al, 2013, p.24). Meanwhile, in the field of ID – a wealth of new knowledge and practice is being cultivated around program evaluation that has remained largely unknown to PM. This knowledge advances the view of evaluation to consider outcomes of programs (Rolstadås, et al., 2014, p. 639). Seeing this gap between what PM has been able to offer program evaluation and what ID has been able
to contribute, this research aims to create more learning opportunities between PM and ID.

Following the selection of constructs, identifying the relationships between said constructs further crystallizes our theoretical stance. In this case we are first interested in observing if and how PM’s contribution to program evaluation has been impactful. This will ask the researchers to observe any links between PM and ID. In doing so, the study can answer the first research question: to what extent has PM contributed to program evaluation. Next the research will explore how impactful ID’s contribution to program evaluation has been. Observing this relationship will allow the research to establish how the ID field has shaped and developed program evaluation, independent or not of PM.

Lastly the research will charter the relationship between current program evaluation practices and PM’s current offer to program evaluation, PM4NGOs, in order to answer the second research question: what can current program evaluation practices inform PM4NGOs? This will be the newest relationship explored in the course of this research as no such link has been explored previously as far can be seen by the researchers conducting this study.

Research Question 1: To what extent has PM contributed to program evaluation?

Research Question 2: How can current program evaluation practices inform PM4NGOs?

1.2.2 Objectives of Study

In reviewing the literature on the evolution of program evaluation within ID, we establish the premise that contributions to program evaluation from PM are insufficient (Besner and Hobbs, 2008; Golini et al., 2014, p.652) when compared to the contributions from the ID field (Liket et al, 2014, p.173). This is the assumed perception given the lack of empirical evidence showing the adoption or use of PM-inspired evaluation practices. This relates to the fact that most PM tools limit themselves to monitoring project implementation and the production of immediate outputs, with little of the assessment focusing on the impact or outcomes for the target population (Bamberger, 2000, p. 96). Nonetheless, the literature also shows that the ID field has been able to fill the vacuum left by PM. By treating outcomes as well as learning as fundamental to program evaluation, there can be a “[increase in] the quality of both the evaluation results and their utilization in program improvement and decision making …” (Liket et al., 2014, p. 173). Based on this, we assert the proposition that: although program evaluation is a function of management; most of the useful contributions are not originating from PM but rather from ID.
In effect, the first part of the research will answer the question: to what extent has PM contributed to program evaluation? The research process and findings will not only bridge two disparate fields but also offer a doorway for future literature to explore PM’s relevance in program evaluation and within the ID field in general.

One of the objectives of this research is to look at the management of ID programs and their evaluation in broader terms so as to analyze the long-term impact, of which, the literature recognizes, little is known (Mitchell, 2014, p. 608). By trying to answer the question of how PM practices have informed and influenced program evaluation in ID, the thesis will try to fulfill an objective of proving or disproving the influence of PM tools and methodologies. This will be seen in terms of improving the existing knowledge about why programs and projects in the ID field have such a high failure rate.

Continuing along this logic, this research will attempt to go beyond considering the application of PM methodologies on the development sector. What has thus far been a one-way conversation between PM towards ID should be promoted as more of a two-way learning channel (Golini et al, 2014, p.657). In light of emerging trends in practice and the great wealth of literature arising from the development evaluation discipline, i.e the area of study dedicated to exploring evaluation in ID; the next assertion is that PM’s non-profit arm, PM4NGOs, could more strategically position itself by paying attention to current program evaluation practices as expressed by INGOs and researchers.

Thus far, PM4NGOs efforts to help better implement ID projects has been carried out as a one-way conversation of PM towards ID, acknowledging the use of standardized, one-size-fits-all, tools specifically in terms of program evaluation. If PM4NGOs looks to inform current program evaluation practices in ID, it could benefit from learning about current trends. Using a largely deductive approach, the study will collect primary data to help fill the knowledge gaps left by literature. Ultimately the findings will be translated into a set of recommendations on how to improve the evaluation chapter of future editions of PM4NGO’s main guide: PMD Pro. These recommendations will address the neglect in the guide’s approach to evaluation by providing a more informed and comprehensive view to program evaluation. In effect the recommendations answer the second question of what can current program evaluation practices inform PM4NGOs? The theoretical proposition we put forth is that PM can upgrade its offer, as articulated through PM4NGOs, to INGOs and the ID sector at large. They can so by drawing lessons from the emerging trends in program evaluation practice and discourse.
Chapter 2. Research Methodology

2.1 Theoretical Principles of Data Collection

Prior to launching into the research, it is worth understanding how the subject will be approached based on the underlying assumptions and implicit worldviews inherent to this study (Collis & Hussey, 2014, p. 43; Morgan & Smircich, 1980, p. 491). Such core assumptions are based on the specific philosophical fields of: ontology, human nature and epistemology (Morgan & Smircich, 1980, p. 491).

Adopting Saunders, et al.’s (2009) notion that the ontological and epistemological assumptions underpin any research strategy (p.108), these orientations can be expected to affect the researchers as well as the object of study. Researchers are likely to have their own biased epistemological and ontological perspectives that have been shaped by their experiences. More importantly, these can stand in contrast to theory and how the scientific community has perceived the subjects of study (Guba & Lincoln, 1994, p. 105). For this reason, it’s important for studies to state their assumptions upfront, thereby acknowledging any potential bias going into the analysis.

2.2 Ontological and Epistemological Orientations of the Research

Ontology is concerned with the nature of reality while epistemology is concerned with how knowledge is obtained in that reality (Morgan & Smircich, 1980, p. 492; Saunders, et.al., 2009, p. 110). These fields move along a spectrum from subjectivism to objectivism. On the subjectivist end the world is interpreted as a projection of the human mind and how it understands its surroundings, in a process of constant codification and creation. Meanwhile on the objective pole the stance is that social phenomena and the interpretation of the world are instead external facts that are beyond the influence of the subject (Bryman & Bell, 2003, p.22)

In considering the subject area of study, program evaluation, as straddling between the fields of PM and ID, we are able to establish the epistemological positions steering the research. PM, for example, has traditionally garnered a more scientific worldview borne of its inherited DNA from the engineering and construction sectors. In effect, its inclinations have always been more positivist than interpretivist. This means that PM typically leans towards a more managerial, technocratic, and instrumental approach (Ika & Hodgson, 2014, p. 1182). Yet it is equally possible to argue that PM ought to be inspected from a constructivist lens since it is also described as “a social construct that evolves with time” (Gauthier and Ika, 2012; Morris, 2011, 2013 cited in Ika & Hodgson, 2014, p.1183).

On the other hand, ID more squarely relies on a more subjective perspective since it has been conceived from a humanistic sector that usually deals with dimensions of human and social development (Ika, et al., 2009, p. 63).
Given that this study establishes itself on largely socially defined constructs and in a humanistic context, it would be most appropriate to adopt a subjectivist worldview (Ika & Hodgson, 2014, p.1183). On this side of the spectrum, it is “necessary to explore the subjective meanings motivating the actions of social actors in order for the researcher to be able to understand these actions” (Saunders, et. al., 2009, p. 110). Moreover, the world is viewed as a continuous process of creation as it is contingent on convention, human perception and social experience (Cars, 2006, p.7).

Although the subjectivist view will be adopted, this does not negate the fact that the fields of study have oscillated between subjectivist and objectivist tendencies during their own philosophical evolutions. For example, PM has increasingly borrowed constructs such as leadership and organizational culture in light of their growing transversal relevance (Ika & Hodgson, 2014, p.1184). Similarly, ID can also be said to have leaned on positivist orientations during its early economics-driven epoch where ID employed rationality and logic (Williams, 2014, p.4). These mixed origins of PM and ID are important to keep in mind as the study pursues its line of investigation on program evaluation.

Nonetheless, the blended epistemological origins of these fields also encapsulate just how much the fields responded to the circumstances of their environment. This gives further reason for the study to adopt an interpretivist paradigm.

2.3 Theoretical Contribution

In uncovering the relationship between PM and ID on the topic of program evaluation, this study takes the stance that the contribution to program evaluation has been disproportionate. For this research proposition to have traction, it should possess both novelty and continuity (Sandberg & Alvesson, 2011, p.23). Continuity is derived from locating the research in established literature. Once continuity has navigated the research through the main debates and concepts, the study will then move to validate the novel characteristic of the proposed research question via critical confrontation. The theoretical contribution is considered a critical confrontation since there is a need to be “critical of a field based on some shortcomings” (Sandberg & Alvesson, 2011, p.38).

In this case, critical confrontation takes the forms of application and neglect gap spotting. Application spotting identifies complementary perspectives that are lacking in an area of research. In this case we have identified PM and its application to the ID field, via PM4NGOs, as the area that could benefit from informing itself on current practices in program evaluation. Once more this is the case given the lack of evidenced impact that PM tools and knowledge currently have on ID programs (Golini, et al., 2014, p. 657). In turn, neglect spotting will address aspects of the research subject that have been overlooked (Sandberg & Alvesson, 2011, p.30). This will manifest as the study’s ability to tease out and accurately capture some of the most salient practices in
program evaluation as communicated by the research and practice community. Among these trends will feature: the rise of outcomes and even participatory approaches, the growing importance of the effectiveness criteria and not exclusively efficiency (Watkins, et al., 2013, p.29), and the improvement in tailoring evaluations to specific programs at hand (Rolstadås, 2014, p. 639).

Ultimately, the gap spotting will be two-fold: observing the gap in PM’s contribution to program evaluation and the reverse: the gap between program evaluation practice and PM’s current offer. Addressing these gaps will necessitate a theoretical framework that is suitable to meet the research objectives of the study and build upon the foundations of the research area.

Upon gaging the knowledge landscape on the proposed subject, the researchers recognized a need to locate the study firmly in the literature by adopting a deductive approach. By nature of being deductive, the framework will locate the relevant concepts and accompanying constructs in the literature in such a way that will help create a credible narrative around the research questions (Saunders et al, 2009, p.61). Establishing clear concepts will also help guide the limited albeit important primary data collection process. In this way, drawing from theory will allow the research to sharpen its theoretical position first while collecting data will fill the gaps the literature cannot answer.

2.4 Research Approach and Method

In adhering to a deductive approach, the researchers will first draw out theories and ideas explaining how the phenomena at hand is observed and then, with the aid of data collection and analysis, verify or reject the assumptions and propositions established by the theory (Saunders, et al., 2009, 61; Ghauri & Grønhaug, 2010, p. 16). By this measure, the critical literature review will extract the most prominent arguments in both fields and complement them with primary data in order to answer the research questions.

Although most interpretivist research prefers to adopt an inductive approach, this research’s line of investigation is found at the intersection between the fields of PM and ID, each with their own wealth of literature. To achieve a cross-sector analysis, and see how these have influenced program evaluation, the study will need to integrate concepts from both fields or at least observe them in parallel. The methodological theory, in this sense, does not dictate any incompatibility in mixing approaches and philosophies. Using mixed methods, as long as these serve a purpose to answer the underlying assumptions posed by the study, can help provide robust findings and help avoid the pitfalls of method bias (Davis, Golicic and Boerstler, cited in Collis & Hussey, 2014, p. 72; Saunders, et al., 2009, 124).
2.5 Data Collection

To bring this cross-fields research into fruition, the research lends itself to a mixed methodological approach, coherent with the underlying assumptions posed by the study (Saunders, et al., 2009, 124). On the one hand, a descriptive approach is adopted to garner a more accurate snapshot of the current state of program evaluation (Saunders et al., 2009, p.140). In this way, the study can better focus on exploring the gaps identified earlier. As such, the study is also strongly exploratory in nature since the research aims to extract the reoccurring trends in program evaluation practice for its inclusion in PM4NGOs. Such an exploratory approach allows the research to start from a broader perspective and progressively narrow down the scope of the study (Saunders, et al., 2009, p.139). That is to say program evaluation, ID and PM will be considered in their broadest sense at first but the research process will reveal their intersecting lines as the evolving logic folds out.

Besides selecting an appropriate methodological approach, research methodologies are also in place to articulate how the data collection will answer the research questions (Collis & Hussey, 2014, p.59). Rather than weighing qualitative and quantitative against one another in a vacuum, the nature of the study should proffer telltale signs of the most suitable methods (Smircich & Morgan, 1980, p. 499). Indeed while this study has taken a decidedly constructivist approach, it does not negate the research from employing mixed methods (Newman, 2000, p.5). A mixed method approach is conceived as pulling from techniques on both the quantitative (content analysis) as much as qualitative (interviews) sides of the spectrum (Long et al., 2000, p.189). As such, the use of mixed methods will also facilitate the triangulation of data findings later on. In other words, it is a practice, for cross-validating more than two distinct methods with the goal of producing comparable data. This allows us to examine the same dimension of a research problem from a variety of lenses (Jick, 1979, p.602). The qualitative leaning offers in-depth look at the subject to explore its complexity (Gephart, 2004, p. 455) while quantitative elements help quantify and bring objective scientific rigor.

The study also acknowledges the pre-dominant use of non-probability sampling given that the sample sizes are not representative of the population, selected by the researchers, and non-generalizable (Saunders et al., 2009, p.213). In particular, the technique of self-selection sampling is the most used in this study since it is well suited to the research questions and needs of a small exploratory study of this nature (Saunders et al., 2009, p.241).

2.5.1 Critical Literature Review

This research’s first purpose will be to gain more clarity on what PM has contributed to program evaluation and second to identify how current practices in ID program
evaluation can inform PM4NGOs. To meet these objectives means critically investigating what the literature has to say.

In order to yield core articles of interest for the critical literature review, a list of keywords was generated following a brainstorming session and preliminary overview of introductory literature (Collis & Hussey, 2014, p.77; Saunders et al., 2009, p. 79). Our key word search targeted terms such as: project management, international development, non-profit management and development evaluation. These preliminary keywords were run through two different university databases: Umeå University and Heriot-Watt University. While no one article explicitly treated all three fields in great depth, each article shed light on one or a few elements of the subject. To push this search further, we also took cues from the reference lists of leading papers. Even so, it quickly became evident that the academic community talking about PM in ID is quite small, with many authors making a recurring appearance in reference lists.

To compliment the theoretical findings in answering the first research question, a critical analysis has been conducted on PM4NGO’s guide: PMD Pro. This guide serves as one of the most telling indicators of PM’s contribution to program evaluation given that it is the only tailored and explicit guide for nonprofits to have come from PM.

### 2.5.2 Semi-Structured Interviews

While the literature review will serve as a first step in the study, the lack of consensus in the literature means that some input from primary data collection is crucial to clarify the bigger picture.

Semi-structured interviews were arranged with various different respondents from INGO practitioners, academics and a PM4NGO representative. The use of semi-structured interviews as a data collection method is motivated by several factors. First, this research aims to have better insight into the world of program evaluation within the sphere of ID. Part of what semi-structured interviews facilitate is an exploration of the relationship between constructs via engaging experts and practitioners on the subject (Collis & Hussey, 2014, p.133; Saunders, et al., 2009, p.322). Indeed, since the study is strongly explorative, the methodology literature endorses the use of semi-structured interviews as a way of uncovering the relationships between the constructs of an exploratory study (Saunders et al., 2009, p.322). Moreover they also serve to reveal meanings, perceptions and interpretations behind much of the constructs and key terms surrounding the research (Cunliffe, 2011, p.659). Second, from listening and internalizing what experts have to say, areas that were previously unexplored by the research might surface, or an idea previously conceptualized from a theoretical standpoint might be confronted from a practical point of view. Consequently, in selecting the respondents, it is crucial that chosen individuals have demonstrated knowledge about the phenomenon, even if those individuals view it from diverse angles (Eisenhardt & Graebner, 2007, p.28).
For the interviews to be fruitful, the researchers have designed specific questions that will touch upon subjects that have been identified previously through theory research (Appendix 2, 3 and 4). Although semi-structured interviews provide some degree of freedom, they also ensure that all relevant themes for the study are covered (Saunders et al., 2009, p. 321; Qu & Dumay, 2011, p. 246).

The first set of key informants selected for the interviews were academics and co-authors Paolo Landoni and Ruggero Golini who have been at the forefront of the discourse on PM’s integration within the ID field. The study deemed them an appropriate choice for interviews given that their work in the subject area is the most cited, albeit among few, actively exploring the current interplay between PM and ID. Engaging Golini in particular was seen as important given that he is the author most cited in this line of research, having published with a number of other academics on various papers. The set of questions laid out in the interview guide aim to complement or extend knowledge that has been documented in the literature around PM in ID.

Semi-structured interviews were also scheduled with six Monitoring & Evaluation (M&E) officers, the industry term for those working in program evaluation, from various London-based INGOs to understand the current approaches and tensions underwriting program evaluation practice. INGOs were identified as those registered non-profit organizations managing international programming. London-based INGOs were selected since London is a known hub for INGO headquarters, which are those most principally charged with the work of program management and evaluation.

Given that both researchers of this study are heavily involved in the non-profit sector, it was possible to pull from collective networks to recruit the correct respondents within INGOs. A pre-made message formally introducing the research was created and a semi-structured interview guide included for full disclosure. This message was circulated to those working within the M&E offices of INGOs. All respondents of the study were M&E officers and therefore the most suitable staff to respond to queries on program evaluation. As they confirmed their interest and availability, the researchers followed up with them by e-mail to confirm and schedule a time to interview at their earliest convenience. All interviews were conducted by both authors together and in English.

The interviews were carried out face-to-face when possible, and otherwise conducted over Skype. The interview guide was constructed to thematically address areas that are not as covered yet pertinent in the literature (Saunders et al., 2009, p. 239), including: approaches and methods used by INGOs, perceptions around who is shaping the methods and approaches, and perceptions of the sector trends. These areas seem particularly relevant to explore since they are precisely the areas of knowledge that could inform the study’s research objective to provide recommendations to PM4NGOs.
The interview guide carried out was the same for every INGO; ensuring continuity in the exploration of these thematic issues (Qu & Dumay, 2011, p. 246).

In addition, an interview was scheduled with a staff member of PM4NGOs to gain insight on how this association has conceived its contribution to program evaluation. Their opinion was deemed relevant to the study since PM4NGOs is a relatively new association and there is limited critical commentary on them in the literature.

Interviews were scheduled to last 45 minutes, with time being evenly distributed to cover each question. In the spirit of semi-structured interviews and in instances were questions were already covered, interviewers would avoid the redundancy of asking those questions and move logically forward to the next intended one. This is one of the benefits of semi-structured interviews, where the interview process can experience an out of order sequence in the line of questions (Qu & Dumay, 2011). Keeping with a more organic flow of conversation, the interviewers are able to follow-up and ask for clarifications around the respondent’s meanings as they naturally arise (Bryman & Bell, 2011, p. 467; Qu & Dumay, 2011, p. 247). In effect, semi-structured interviews afford some degree of freedom while guaranteeing that all relevant themes are covered (Saunders et al., 2009, p. 321).

All interviews were recorded (with permission of the respondents) and transcribed to allow researchers accuracy and facility in data presentation and analysis.

<table>
<thead>
<tr>
<th>Interviewee</th>
<th>Length of Interview</th>
</tr>
</thead>
<tbody>
<tr>
<td>Plan UK</td>
<td>49 min. 32 sec.</td>
</tr>
<tr>
<td>Restless Development</td>
<td>55 min. 10 sec.</td>
</tr>
<tr>
<td>Youth Business International</td>
<td>43 min. 53 sec.</td>
</tr>
<tr>
<td>War Child</td>
<td>47 min. 28 sec.</td>
</tr>
<tr>
<td>Action Aid</td>
<td>58 min. 26 sec.</td>
</tr>
<tr>
<td>Oxfam GB</td>
<td>50 min. 57 sec.</td>
</tr>
<tr>
<td>Ruggero Golini</td>
<td>51 min. 27 sec.</td>
</tr>
</tbody>
</table>

*Table 1: Organizations and people interviewed and length of each interview (Source: Own Creation)*

### 2.5.2.1 Limitations of Interviews

Although both academics were contacted for individual interviews, only the interview with Professor Golini was carried out, as Professor Landoni was unavailable. Landoni commented to the researchers that as he and Golini worked together in publishing the research, the merit of uniquely interviewing Golini would be just as high. As Professor Golini is based at the Politecnico di Milano, the interview was conducted over Skype.
Meanwhile a small self-selected sample of six interviews was completed with the London-based INGOs of War Child, Youth Business International, Oxfam UK, Plan UK, Restless Development and Action Aid. The sample was focused on INGOs that shared specific characteristics. In this case there were two criteria that permeated the selection: location and size of the organization. The INGOs had to be mid-sized to large INGOs running several different programs across the ID landscape.

Although researchers intended to reach ten, lack of time and general time of the year for organizations made scheduling interviews particularly challenging and reduced the sample to six. Some of the reasons for their limited availability included: inconvenience of end-of-year season, and bad timing due to organizational restructuring. The researchers own time constraints also limited the number of interviews that could be run. Even so, it is common practice to make use of smaller samples with a technique like interviewing as it allows for more in-depth investigation (LaForest, 2009, p.2).

In terms of the interview with PM4NGOs, the researchers were connected with Edson Marinho from PM4NGOs. However due to lack of availability for an interview on his side, he requested instead to receive questions and respond electronically. In lieu, a questionnaire with the same questions from the interview guide was issued by email. The limitation here is the limited ability for researchers to follow up extensively with his answers. The aims of the questions were to afford the researchers a more nuanced view of how the association has conceived its program evaluation offer. Regrettably the answers provided did not yield a great deal of new insight, as they were too broad and offered information already available in the PMD Pro guide or on the website. Rather what his responses provided instead was a affirmation of the above mentioned PM4NGOs sources and extended some of the factual knowledge. The results of this questionnaire are made available in the appendices for reference and full disclosure (Appendix 3). The researchers of this study opted for inclusion of those questionnaire results in the appendices rather than data presentation since most data collected reasserts what has already been covered by the guide and website, as best seen in section 3.8 titled ‘PMD Pro: A Critical Analysis of PM4NGOs Contribution to Program Evaluation’. Future studies could more closely scrutinize the role of PM4NGOs with several and more extensive interviews with various personal, but given the time limitations and the distance of the office (located in Brazil), this was not possible at this time.

2.5.3 Content Analysis

Content analysis is envisioned as a process of systematic and quantitative description of content (Bryman & Bell, 2003, p. 194). The study employed this technique with the purpose of understanding what is currently being communicated in calls for evaluation. These manifest in the form of Terms of References (ToRs) released by either donors (funders of programs) or INGOs interested in commissioning an external evaluation team to complete the evaluation of their program. By spotting the trends in ToRs, the
study should be able to corroborate these findings with those extracted from interviews and literature review; and as such - comment on the current approaches to evaluation.

The ToRs used for this content analysis were accessed through the Pelican Initiative - an internal network of practitioners involved in program evaluation within the ID field (Pelican Initiative, 2005). Although the group has closed members, the researchers have membership and consequently were able to access ToRs circulated by INGOs and donors. The twenty ToRs that were ultimately selected were the most recent to have been published in the group, the full listing of which can be found in the Appendices (Appendix 5 and 6).

The main advantage of running a content analysis is it lends itself to systematically code the qualitative data (evaluation criteria frequently reoccurring in ToRs) into numerical data (frequency with which criteria recur) (Collis & Hussey, 2014, p.166). Being able to breakdown the current calls for program evaluation in this way will encapsulate what is happening in practice. In effect, it will complement findings from the literature review and interviews to inform researchers on what current practices exist in program evaluation that are worth noting for PM4NGOs.

2.5.3.1 Limitations of Content Analysis

In selecting the sample size, the researchers took into consideration the trade off between efficiency and effectiveness. That is - that the sample should be large enough to yield data of some significance to the research (effective), but for these findings to be achieved using the smallest data set (efficient) (Wang & Riffe, 2010). With this tension in mind, and understanding that “[no] sampling guidelines exist for researchers to select an adequately effective sample in examining content [...]” (Wang & Rife, 2010), the researchers settled on a sample of twenty ToRs. This size of sample is enough to allow for in-depth analysis while also being large enough to serve as a representative sample. The principal behind this is that after this amount, no new data could be extracted, i.e a point of theoretical saturation (Fugard & Potts, 2015, p. 671).

The selection of sample size is also subject to time and information constraints. Collis and Hussey (2014) state, “if the data is manageable and you have sufficient time, you can analyze all the data” (p. 165). However, in this case there are two fundamental limitations: 1) amount of data readily available 2) researchers’ time. The first point refers to the fact that ToRs come out sporadically and are generally removed after a period resulting in a limited number of available ToRs. With regards to the second point, the researchers are constrained by time, having only two and a half months to conduct, analyze and present the research.

Another caveat worth recognizing is that the sample has some organizations, like Oxfam, featured more than once as they release a higher volume of ToRs to the Pelican Initiative. Although it is not expected to skew the data set in any major way, it is worth
considering that this organization may go about creating their ToRs with a similar approach and language thus leading to unintended redundancy. The same comments can be said about ToRs stemming from UN bodies that are featured in the sample. This being said, each call for evaluation is done for a unique program in a different country; each of these programs adhering to a particular set of needs. For this reason the researchers did not feel that data would be skewed in any significant way and were able to proceed with the content analysis.

2.5.4 Triangulation

The investigation ultimately used triangulation to attest that the theory extracted from the literature and the data collected through the content analysis and semi-structured interviews show comparable and similar results. Triangulation, as stated by Saunders, et al. (2009) “refers to the use of different data collection techniques within one study in order to ensure that the data are telling you what you think they are telling you” (p. 146).

Effectively by triangulating findings from the literature review with the content analysis and interviews, it should bring to light the most noteworthy practices in program evaluation. Ultimately, this study will attempt to produce a small set of recommendations to PM4NGOs on how their guide ‘PMD Pro’ can be updated to include a more relevant program evaluation chapter.

2.5.5 Ethical Considerations

The pursuit of this investigation does not come without a few worthy ethical considerations. Ethics is the way through which researchers can negotiate access to information, organizations and people (Saunders et al, 2009, p.160). This applies to the study given that INGO representatives, researchers and practitioners are involved.

For INGO representatives, the question of ethics is an important one since they are highly politicized bodies where the relationship between negative publicity and risk of losing funds is delicate (Cooley & Ron, 2002, p.6; David, 2006 cited in Carman, 2010, p.259). Consequently, to contact INGO M&E officers, the questions were made available up front and anonymity guaranteed. That is to say, names of individuals will not be provided. As will be seen in this research, while the full list of organizations interviewed is disclosed, the researchers have coded them so that references are attributed to an organization, not an individual. Indeed the guiding principal for ethics is that “the research design should not subject those you are researching (the research population) to embarrassment, harm or any other material disadvantage” (Saunders et al, 2009, p.160).

As for the interview with Professor Ruggero Golini and questionnaire issued to Edson Marinho of PM4NGOs, they provided permission to participate and be named in the
study. The risk of poor exposure is less hazardous for these individuals operating in less volatile contexts.
Chapter 3. Theoretical Framework

3.1 Concepts Definition

The upcoming literature review will relate the concepts to the purposes of the research. To better set the scene, we begin by clarifying the two fields of interest: ID and PM.

To comprehend ID, it is useful to envision the sector as a constellation of actors working together towards alleviating poverty and improving quality of life in developing countries (Mitchell, 2012, p.605). Among its principle characters are donors (the funders of ID programs), INGOs (the implementing agents) and beneficiaries (the recipients of the program service) (Hirschman, 1967; Ika, 2012; Khang and Moe, 2008 cited in Ika & Hodgson, 2014, p.1186).

Quite differently, PM is conceptualized as a field aimed at providing a body of knowledge on planning, organizing and managing resources to bring success to projects, programs and portfolios (Turner, 2009, p.1). It’s important to note that the discipline is widely referred to as ‘Project Management’ and many of its tools refer to ‘projects” since it’s the most basic unit of analysis. Nonetheless, our research adopts the unit of analysis of ‘programs’ – “a group of projects which contribute to a common, higher order objective” (Turner, 2009, p.324; Culligan et al., 2011, p.9). This choice is motivated by the fact that management through programs rather than projects is far more commonplace in the ID sector today (Holvoet and Renard, 2006, p.67).

To explore the role PM has played within ID, this study adopts Ika and Hodgson’s (2014, p.1189) three paradigm shifts: conventional, people-centered and critical (see table 1). These shifts are crucial to consider, as they will draw attention to the evolution of PM in ID. In exploring these paradigm shifts, two things should become evident: one, that the fields of PM and ID have always been inextricably linked and two, that the microevolutions within PM in ID parallel similar developments in program evaluation.

After establishing the two fields at the bedrock of this research, the literature review will tackle the overarching concept framing our study: program evaluation. Upon scanning the evaluation literature in the ID field, it is evident that authors do not agree on any one overarching definition for program evaluation (Boulmetis & Dutwin, 2005; Gray & Associates, 1998; Hatry, 1999; Mullen & Magnabosco, 1997; United Way of America, 1996 & W. K. Kellogg Foundation cited in Carman, 2010, p.264). This is namely true because the notion of program evaluation is associated with a myriad of different viewpoints and approaches (Holvoet and Renard, 2006, p.67) as well as different paradigms. As such, there is no piercing consensus (Snibbe 2006, cited in Liket et al, 2014, p.173).

One of the principle reasons for this is that managers working with INGOs have different conceptions of evaluation - ranging from treating it as a highly systematic process to an informal reporting of how the program has been doing (Murray 2005, in
Carman, 2007, p.62). It is also important to note that this study refers to the field of evaluation within the particular context of ID and not simply ‘evaluation’ since the latter is its own broad discipline entirely (Benjamin, 2012, pg. 440). Meanwhile within the PM literature, a poor understanding of program evaluation has led to amalgamate certain functions, like monitoring and control, with evaluation when in fact they are each quite distinctive, albeit complementary, in purpose (Bamberger, 2000, p. 96; Makarova & Sokolova, 2014, p. 77).

Nonetheless, this study finds a point of departure for the concept in the Handbook of Practical Program Evaluation, describing program evaluation as “the systematic assessment of program results and, to the extent feasible, the systematic assessment of the extent to which the program caused those results” (Newcomer et al., 2004, cited in Carman, 2007, p.62). By this definition, understanding results and the process by which results were generated by the program are the basis of an evaluation. However this notion of evaluation will not be taken strictly as this definition but rather observed across three distinct paradigms.

The research observes program evaluation in terms of paradigm shifts as a way of recognizing the important evolution that it has undergone and to draw out its most salient features to date. Although no one author is able to offer an updated timeline of program evaluation’s evolution, scouring the literature we were able to tease out some noticeable shifts in thinking. Three paradigms in particular have emerged: Traditional Evaluation, Outcomes Evaluation and Participatory Evaluation as characterized by authors Rebien (1999), Buckmaster (1999) and Suarez-Herrera et al (2009) respectively (Appendix 1). As can be evidenced by the table, each paradigm possesses its own distinctive features that move evaluation approaches on a spectrum from outdated conventional thinking and practice (traditional evaluation) to contemporary but not yet fully adopted theory and practice (participatory). These three paradigms will guide the analysis on the contribution from the ID field to program evaluation. By identifying how program evaluation practice has changed, the shifts in paradigms should indicate to PM how it should best adjust its offer to stay up to date.

In effect, an integrated framework is proposed as a way of analyzing paradigm shifts for PM in ID programs in tandem with the paradigm shifts for program evaluation practices.
Figure 1: Paradigms Within Program Evaluation and Project Management in the Development Sector (Ika & Hodgson, 2014; Rebien, 1997; Buckmaster, 1999; Suarez-Herrera, et al., 2009)

A critical review of the literature on PM in ID and program evaluation is the first step this study takes in answering the proposed research questions. To build a credible narrative, the research will first establish and characterize the context: the ID field. This will be followed by a critical scrutiny of paradigm shifts observed in PM & ID as well as in program evaluation.

3.2 Assumptions and Limitations

Certainly no research comes without its inherent assumptions (Collis & Hussey, 2014, p.43). Upon scrutinizing the literature and coming to understand program evaluation as a field with little consensus, it led to the assumption that no research has been done to categorize and capture program evaluation in theory or in practice. Consequently, in pursuing our theoretical proposition, in an under researched niche area, we expect that our own limited understanding, and bias, will affect the way reality is observed. This once more reaffirms why interpretivism is assumed as the most favorable epistemological stance for this study.

Another noteworthy assumption is that PM in ID, and program evaluation have passed through the three parallel paradigm shifts, as presented in the theoretical framework earlier. Although the paradigm shifts are squarely supported by evidence in the literature; i.e pulling from Ika and Hodgson’s propositions (Ika Hodgson, 2014, p.1189) and by extracting trends in development evaluation literature; there is no wide consensus to empirically validate these observed paradigm shifts. This is most true for the development evaluation discipline that has been characterized by a lack of organization and coherence. Thus, it is important that future research further investigates some of the propositions that will be made here.

In terms of limitations, there are a few worth mentioning. First, the journals available on the subject of study are mainly concentrated around academia from the United States, Canada and United Kingdom. Absent from the discourse is thus the perspective of local offices in target countries where the data collection that inputs into program evaluations actually takes place. This asymmetry is not just symbolic of the imbalance
of perspectives in the development evaluation discourse but also serves as a reminder that there is much more research needed in this area. However, there is no scope in this research to investigate other countries as “proper comparisons among countries around the world [prove] difficult due to methodological issues, conceptual developments, and empirical focus” (Hvenmark & Larsson, 2012, p.60). Even so, not being able to actually assess the perspective of evaluation from country offices limits the breadth and understanding of this research.

A further specification made for this research was the selection of INGOs headquartered in London. It made particular sense to study headquarters since they tend to oversee the larger strategic performance of the organization (Crawford & Bryce, 2002, p.367), including oversight over program evaluation. Not only that but the most modern management practices are taking place in headquarters, not in country offices (Barr & Fafchamps, 2006; Barr, Fafchamps, & Owens, 2005; Burger & Owens, 2010; Watkins, Swidler, & Hannan, 2012 cited in Marshall and Suárez, 2013, p.1034). Moreover, INGOs are often those that shape trends in the sector (Marshall and Suárez, 2013, p.1045) and for this reason as well they would be the most relevant to study. Although the organizations sampled have different missions, they would still be deemed a reliable sample since they share at least the common feature of having programs in a number of developing countries.

Ultimately taking into account the ‘who, where, when’ of the theoretical proposition by recognizing these assumptions and limitations is fundamental to determining the study’s generalizability (Whetten, 1989, p.492). In this case, the research is venturing into a new unchartered area and will be bridging three areas that are seldom discussed together in the literature. For these reasons and those mentioned above, the study’s findings will not be generalizable. Rather it should serve as a diving board for future researchers to investigate and generate empirical data that can be tested.

### 3.3 International Development

ID, as an academic area of study and a field of practice, serves as the backdrop of our study. As a result, it is important to understand where development comes from, how it was formed and what is the nature of the field today. Although it may seem unlikely, the origin story of ID can be traced back to Adam Smith, who advocated for the careful balance between government, economic policy and modernity in terms of nation-states (Williams, 2014, p.3). Although this idea of ID has been around since the latter half of the 19th century, it is not until the mid 20th century, in the post war era, that the field of development took on its current form.

ID has been described as a field dedicated to transformational change by supporting the socio-economic growth of communities in the Global South, otherwise referred to as ‘developing countries’ (Ika and Hodgson, 2014, p.1182). However in its beginnings, it was the ‘economic’ growth aspect of development that was pushed more ardently,
where the increase in wealth of a country was equated to prosperity (Harriss, 2014, p. 2).

To be sure that these ideals and standards were replicated and the increase of wealth could be monitored, development used the backing of the field of economics based on its expertise and “scientific knowledge” (Williams, 2014, p.3). As proof of this, positivist tools were developed to explain growth; including specific standards for measuring development, exemplified by the creation of indexes such as the gross domestic product (GDP) to gage economic productivity of a country (Harriss, 2014, p. 1).

Yet this idea of change through socio-economic growth in ID was not only tied to the field of economics. ID was also laced with underlying strategic geo-political incentives embedded in the discourse of power and domination (Ika & Hodgson, 2014, p.1888). In deed, the early days of development came about during the Cold War era. U.S. policy-makers recognized a strategic need to integrate countries that they considered developing, into a multilateral, political and economic new world order they were trying to build (Williams, 2014, p.3). By expanding the concern for growth to developing countries, there was subsequently an excuse to export the western worldview. However in doing so, this created incongruence between western ideals and imposed best practices and realities in developing countries.

Since then, the notion of development has morphed into something different. The ideas and ideologies that ushered in development no longer hold water. Among the outspoken critiques, Amartya Sen (1999) argued that production and commodities, in term of economic growth, are only of value in terms of what they allow people to do (p. 231). The weight of those words heralded a change so that development was not only tied to the wealth of countries but to power and freedom behind the choice of individuals. Sen’s people-centered vision inaugurated a new wave of thinking that has become the staple of ideals in the field today. Even so, ID is also known to be a field where the romanticized and idyllic notions in theory have a hard time translating into practice. Although the field of ID is moving away from its early ideals and concepts of what change and growth mean, its systems and politicized dynamics still insist that systematic change can be captured as measurable progress (Harriss, 2014, p.4); in effect showing traces of its historical conception.

3.4 PM in the field of ID

In light of the ID’s genesis, it is not hard to imagine a time when PM and ID shared philosophical origins. Ika and Hodgson (2014) claim that the use of PM in the ID sector can be traced back to the 1950’s (p.1182). At the time, these fields shared a common appreciation for instrumental rationality, objectivity, reductionism and expectations of universal validity (Ika & Hodgson, 2014, p. 1184; Golini & Landoni, 2014, p. 122; Ika,
et al., 2009, p. 62). However, to understand how PM in ID has become an offshoot from traditional PM it is worth distinguishing their differences, similarities and, how theory and practice have tried to reconcile PM and ID.

PM can be traced back to the middle of the 20th century (Morris, 2011, as cited in Ika & Hodgson, 2014, p. 1182). It rose to prominence with the technological successes of the aerospace and the military-defense industry of the 1950’s and 60’s (Ika & Hodgson, 2014, p. 1183) and effectively established itself within the WETT industries (water, energy, transport, telecommunications) (Turner, 2009, p. 2). To this day, PM’s identity shares a strong affinity to these types of ‘hard’ programs (Golini & Landoni, 2014, p. 122). They are referred to as ‘hard’, given that these programs possess controllable and mechanistic processes, where the fixed nature of program inputs translates to low uncertainty (Crawford and Pollack, 2004, p. 646). It has also meant that the relationship between input and output is considered logical (Crawford & Pollack, 2004, p. 649) and programs generally feature few stakeholders (Golini & Landoni, 2014, p. 122).

Differently, ID programs are characterized by being less fixated on logical relationships between inputs and outputs given the intangible nature of program goals (Corti & Landoni, 2011, p. 45). Indeed, ID distinguishes itself in that it has evolved away from hard ‘service delivery’ style programs towards ‘soft’ programs with more amorphous goals. Such objectives concern areas such as poverty alleviation and improvement of living standards, environment and basic human rights protection, among others (Ahsan & Gunwan, 2010, p. 68). This notion concurs with Crawford and Pollack’s (2004) description of ‘soft’ projects having broader political, communal, and social impacts (p. 645).

ID programs are also distinct in that their stakeholders are more diverse and more numerous than in traditional WETT programs (Turner, 2009, p.2). Moreover, the riskier disposition of the environment creates higher uncertainty since inputs might have to change in response to resource scarcity or the cultural/political/social context (Golini & Landoni, 2014, p. 652). In this respect, the difference between traditional programs and ID programs holds merit.

However, the ill fit of PM for ID programs is not just a simple differentiation in the nature of programs and context. It also relates to the inappropriateness of traditional tools, explaining why there has not been widespread use and adoption (Watkins et al., 2013, p.29). As can be inferred from the narrative thus far, traditional PM tools were crafted with traditional business sectors and ‘hard’ programs in mind, not the ‘soft’ programs of ID. Of the few studies conducted, ID is witnessed as having a very minimal uptake of PM techniques, with virtually none cited for program evaluation in particular (Golini et al, 2014, p.652). With no evidence to give credibility to PM’s relevance in ID programs, ID program managers have remained unconvinced by the use of PM tools (Golini and Landoni, 2014, p.652; Ika & Hodgson, 2014; Diallo, Ika & Thuillier, 2010).
Researchers state that the problem of tool usage can be twofold. On the one hand, conventional tools can serve a practical purpose where, for example, greater use of these tools tends to increase program success since ID programs notoriously lack systematic planning methodologies (Golini & Landoni, 2014, p.10). Meanwhile other authors purport that the problem is in the tools themselves, since they are inadequate to quantify and qualify the reach and impact of ID programs (Ika & Hodgson, 2014, p. 1186). Even when the practical side of an ID program is to provide a set of deliverables like a road or a school, the goal must meet a specific value for the beneficiaries; it must improve their overall social condition. In this regard ‘hard’ program tools can be a means to an end but hardly ever the end in itself (Ika & Hodgson, 2014, p. 1185; Golini & Landoni, 2014, p.122).

Perhaps the most important reflection to underscore here is that although PM tends to be uni-directionally applied to others sectors, this should not negate that those others sectors cannot inform PM. Indeed, researchers who have investigated the integration of PM in the ID field have insisted that ID may have insights to offer PM (Golini et al., 2014, p.651). More importantly, the desire for PM to learn and make itself useful, as demonstrated by the mere creation of PM4NGOs, underscores why it is important to conduct this study.

In fact, ID has been quick to react to its own management needs by creating or adapting existing tools to better-fit development work. Several development cooperation agencies have created their own tools and frameworks for ID programs (Golini et al., 2014, p.650). USAID, CIDA and the World Bank are among the multilateral and bilateral agencies that have stepped up to the challenge of creating more appropriate tools for PM in development, including the Logical Framework Approach (LFA) – a framework that helps plan programs by outlining the relationship between inputs, activities, outputs and results (Golini et al., 2014, p.650). Inlaid in these actions is the recognition that traditional tools fall short in meeting the needs of ID.

These developments, however, did not deter PM associations from trying to become more relevant, as illustrated with the launch of PM4NGOs in 2010 (Golini, et al., 2014, p.652) – an association aimed to bring PM knowledge to international and national NGOs. Under their preveue, PM4NGOs produced their own smaller body of knowledge and specific certification: PMD Pro. Although PM4NGOs and PMD Pro have been tailored to include the specific characteristics that portray the peculiarities found in ID, the tools prescribed echo much of what is found within the premier and traditional associations of PM like the Project Management Institute and the International Project Management Association (Golini, et al., 2014, p.651).

Having stated the difference between ID and PM, as well as what has been done to reconcile them; it is also important to highlight the points where both intersect. ID and PM may differ in the type of problems they are trying to resolve however they share a
belief that projects and programs are the way to go about addressing those problems. In fact, the lion’s share of development assistance provided by multilateral organizations and governments is realized through projects and programs (Corti & Landoni, 2011, p. 45). Not an uncommon feature, many fields whose external environment is dynamic, complex and uncertain, tend towards project-based forms or organizing (DeFillipi & Leher, 2011, p. 64). This kind of environment certainly concurs with that observed in ID (Golini and Landoni, 2014, p.124). As such, ID continues to be, to this day, a field that is pre-dominantly project-based or rather program-based.

What the narrative should highlight thus far is that PM and ID have been interwoven at different points in their story but some critical differences have brought them apart. In highlighting them here, this review can bring together relevant aspects of the ID and PM discourses that have shown up scattered in the literature (Ahsan & Gunwan, 2010, p. 68-69; Ika & Hodgson, 2014, p. 1184).

### 3.5 Growth and evolution of PM in ID

The literature has stated that PM grew out of a conventional school of thought – one that championed the universal applicability of tools, methods and cycles. ID was perceived as a sector that seems fertile for adopting PM. Yet the unavoidable differences characterizing ID programs necessitated that PM customize itself to fit development (Ika & Hodgson, 2014, p. 1183). In learning to articulate its contribution to ID, PM has been described to have witnessed three paradigm shifts: 1) traditional or ‘blueprint’ approach, 2) people-oriented or ‘contingent’ approach and 3) critical approach. Specific traits of these shifts are highlighted in table 1.

The first instances of PM in ID were captured in the post second World War reconstruction programs of Europe. At the time, programs adhered to a traditional ‘blueprint’ approach to PM where efficiency was paramount. Indeed the development industry of that time was more fixated on solving technical issues (i.e. re-build roads, clear war debris). In other words rational and controlled approaches were favored as the best way to structure and deliver on ID program goals and objectives (Ika & Hodgson, 2014, p.1187).

This meant that the focus was more on implementation rather than generating impact and outcome of a program. One of the chief examples of this blue print approach within ID is Baum’s Project Cycle approach. Baum identified that any given project or program will go through a series of steps or phases that will connect the beginning to the end, and that these steps should be aimed at providing a well-defined structure and direction to a project’s/program’s activities to achieve its said goals and objectives (Landoni & Corti, 2011, p. 46). This line of thinking supports the definition of program evaluation, described earlier as “the systematic assessment of program results and, to
Table 2: Key Characteristics of Paradigm Shifts Within Project Management in the Development Sector (Ika & Hodgson, 2014, pp. 1183-1189)

<table>
<thead>
<tr>
<th>Key Characteristics</th>
<th>Traditional / Blue Print Approach</th>
<th>Contingent / People – Centred Approach</th>
<th>Critical Approach</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Projects are normative and prescriptive. They are based on the belief that that International Development only poses a technical and managerial problem. Control and structure are the best ways to solve these problems.</strong></td>
<td>This approach reacted to the traditional blueprint approach. It emphasizes putting people first and looks to empower them to solve the problem. In this case the approach is more flexible and fluid, looks at the participation of the different stakeholders (specifically beneficiaries), and preaches that there are no real recipes to development, but that each situation is unique and the solutions to problems will need to be assessed under their particular circumstances.</td>
<td>This paradigm challenges the two former paradigms, arguing that the underlying power dynamics evident must be overturned and rethought and that PM could be a vehicle to accomplish this. Participation could be taken much further. However more than this, this paradigm does little to propose an alternative.</td>
<td></td>
</tr>
<tr>
<td><strong>Reductionism, application of mechanistic rationality</strong></td>
<td>Participation, learning, empowerment, people first</td>
<td>Challenging underlying power dynamics and promoting more honest and equal participation.</td>
<td></td>
</tr>
<tr>
<td><strong>Iron Triangle: Time, Cost and Quality</strong></td>
<td>Participation is key: The involvement of beneficiaries and satisfaction of the majority of stakeholders is the measure of success</td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td><strong>Cost-Benefit Analysis, Logical Framework, Internal Rate of Return</strong></td>
<td>Action planning workshops, stakeholder consultation, participatory rural appraisal</td>
<td>N/A</td>
<td></td>
</tr>
</tbody>
</table>

the extent feasible, the systematic assessment of the extent to which the program caused those results” (Newcomer et al., 2004, cited in Carman, 2007, p.62). The keywords here being: results and demonstration of causal links.

However, as the times changed, the needs of the ID field shifted towards education, health, civil society and securing livelihoods. By the time the 1980s rolled around, a more people-centered development was being ushered in. More eclectic, contingent or middle-range approaches made their debut, criticizing the universality promoted under the traditional paradigm (Ika & Hodgson, 2014, p. 1188). In exchange for the blueprint views, new kinds of programs with more intangible and amorphous goals manifested including aspects like training entrepreneurs and empowering youth leaders.

Under this people-centered paradigm where more contingent approaches where commonplace, two important conclusions can be drawn from the theory: First, the movement recognizes that ID programs have a greater a plurality of stakeholders whose interests need to be balanced even though they may in fact conflict with one another (Golini & Landoni, 2014, p.124). In response, the ID field began to introduce different tools that valued more participatory approaches (whose specific forms are numerous but not within the scope of this study to discuss). The idea behind these approaches was to put people first by empowering them (Ika & Hodgson, 2014, p. 1188) to be more active agents in the execution of programs. Second, the people-centered paradigm embraces the notion of programs having to adapt to political, cultural, historical dimensions within organizations (Ika & Hodgson, 2014, p. 1186), because no two problems are exactly the same. This means that there is no structured, prescriptive, rational approach because the challenges of development are not well-structured problems that can be
thought through with a formula.

Although these two previous approaches contrast each other in the way they look for a viable solution to development problems, the critical school of thought tries to stir academics and practitioners to be more provocative. This paradigm asserts that there is a paradox in the way development and its implementation through programs or projects is done (Ika and Hodgson, 2014, p.1191). It asserts that neither the traditional blueprint approach, nor the contingent, people centered approach have mitigated the high instances of program failures (Ahsan & Gunwan, 2010, p. 69). Thus the critical approach claims that development is a struggle between power and knowledge that is currently imposed with a top down perspective. This means that the ideas behind development itself actually have implicit structures that make the receiving party (less developed countries) submissive towards those imposing their views (aid donors, developed westernized world) (Harriss, 2014, p.3). However in exchange for its criticism, the critical paradigm falls short in providing any credible alternative. Instead, it has been able to set the stage for a broader discussion challenging the development industry’s current form.

3.6 The Rise of Program Evaluation

The high rates of project failure, at an operational and strategic level, in the development field has not only had the effect of bringing the sector under scrutiny but also “prompted a call for better management, accounting and impact assessment systems for non-governmental organizations (NGOs) in order to enhance their ‘social impact’” (Golini & Landoni, 2014, p. 12). This call to action propelled the development evaluation sub-sector forward. Adopting Crawford and Bryce’s denotation, “‘evaluation’ is the periodic (typically at mid-term and end-of-project) examination of the project in its entirety” (Crawford & Bryce, 2002, p.368). More than just a glorified term, the field began to introduce evaluation within the project life cycle officially since 1978 (Landoni & Corti, 2011, p. 46).

Since then evaluation has become one of the most overused albeit least understood terms of the ID field. Ubiquitously referred to with its counterpart ‘monitoring’ to create the household term ‘Monitoring and Evaluation’ (M&E) (Crawford & Bryce, 2002, p.366), the essence of evaluation today is diluted (Marshall and Suárez, 2013, p.1035). At its simplest, monitoring serves a reporting function while evaluation corresponds to a slew of purposes that usually answer the question: did the development intervention lead to a change in the intended goal (Walker, 2015, p.77)? To unpack evaluation, there needs to be an understanding of its origins and its evolution to its current form.

The challenge the literature poses in accomplishing this task stems from a lack of coherence and organization in program evaluation thinking (Liket et al, 2014, p.173). More than just a noteworthy detail, this characteristic is emblematic of the convoluted
nature of evaluation as a practice. In trying to identify the source of the vagueness, a few plausible culprits can be singled out. For one, each donor (funders of programs) carries different expectations of how evaluations should be run and subsequently impose different degrees of control over INGOs (Bamberger, 2000, p.96). In an environment where donors have competing visions, and organizations have a host of different programs to evaluate - this results in a dizzying array of options for program evaluation to chose from (Holvoet and Renard, 2006, p.67). More pertinently, it puts the pressure on INGOs to juggle the particular evaluation demands of each one of its donors along with its own internal organizational needs (Lennie and Tacchi, 2014 cited in Lennie et al., 2015, p. 330). Even if INGOs decide to commission their evaluations to external consultants, a growing practice to prove accountability (Snyder & Doan, 1995, p.150), then the same expectations would still have to be balanced, with the added layer of selecting suitable external evaluators.

However the lack of coherence can also be chalked up to the lack of a golden standard in evaluation. Indeed while there have been frameworks that have experienced more widespread application, such as the LFA mentioned earlier on (Hermano et al, 2013, p.22), these tools have more to do with capturing a summative snapshot (ultimate results) (Crawford and Bryce, 2002, p.368) and less to do with the learning and program improvement goals of evaluation (Christensen & Ebrahim, 2006; Cutt & Murray, 2000; Ebrahim, 2005 cited in Benjamin, 2012, p.433). As will be discussed in the next section, these frameworks ratify views compliant with a traditional paradigm of evaluation when in fact the discourse and INGOs are trying to embrace the learning dimension of evaluation and not simply accountability (Crawford and Bryce, 2002, p.363).

3.7 Program Evaluation and PM: Bridging the Gap

Based on the premise that PM’s offer to program evaluation is insufficient; PM could potentially gain insight from program evaluation practice and discourse. However in order to gage how feasible this is, it is first important to establish what program evaluation has to offer. Herein lies the challenge. The increased spotlight on this sub field has led to the unchecked proliferation of evaluation approaches, tools and jargon (Holvoet and Renard, 2006, p.67). The countless options available to non-profit program managers coupled with top-down pressures from donors to use their preferred evaluation approaches (Snyder and Doan, 1995, p.145) means that program evaluation looks different in every organization (Benjamin, 2008; Carman, 2009; LeRoux & Wright, 2010 cited in Marshall and Suárez, 2013, p.1035). There is thus a need to pour into the literature, filter out the trends and draw out the characteristics that could inform PM’s contribution to program evaluation in INGOs.

Even at the level of evaluation terminology, the Development Evaluation discipline has produced an undeterminable amount with even more nuanced terms seemingly on the rise (Snibbe, 2006 cited in Liket et al, 2014, p.173). Part of this can be understood by
the paradigm shifts taking place in the field, each bringing with them changes to the terminology, tools, approaches and ethos. Unlike PM, the paradigm shifts in program evaluation have blurred lines and it is not entirely clear which of these paradigms INGOs ascribe to. However as established in the theoretical framework earlier, the literature does suggest that there are loosely three paradigms that are most notable: traditional evaluation, outcomes evaluation and participatory evaluation. From left to right, the paradigm shifts reflect a movement from conventional conceptions of program evaluation to more innovative interpretations. Table 2 included in the appendices reflects these shifts (Appendix 1).

The first paradigm, titled ‘traditional’ evaluation is perhaps one of the most elusive to capture in the literature. One of the most probable rationales for this is that the term ‘evaluation’ was not yet adopted in the pilot era of ID programs where instead performance measurement was discussed (Buckmaster, 1999, p.187). More than just a change in lexicon, the reference to ‘performance measurement’ reflected the fact that ID programs adopted PM practices of the 1950’s with virtually no tailoring to the ID/non-profit context (Ika and Hodgson, 2014, p.1182). Since then, the language of performance measurement has been incorporated as a critical aspect within a wider vision of ‘evaluation’. Nonetheless, the traditional evaluation paradigm still reflects much of what PM continues to contribute to ID today: a rational and logical decision-making process (Rebien, 1997, p.454) that is often, although not always, ill fitted for the ID sector.

Indeed, performance measurement, as an approach to evaluation, was upheld as how organizations ensure that inputs led to outputs (Rebien, 1997, p.453). By understanding the former as resources fed into the program and the latter as products produced as a result of the input (Buckmaster, 1999, p.188), it becomes evident that project efficiency (‘doing things right’) was of utmost concern under this paradigm (Diallo & Thuillier, 2004, p.21). Although PM may have adopted some of the modern evaluation language, the thinking and contributions to ID programs remain stuck in the traditional ‘performance measurement’ evaluation era (Liket et al, 2014, p.183). In making the link back to PM, much of the nature of this evaluation era echoes the thinking embedded within the conventional PM paradigm.

The logical framework and results based management (RBM) models are among the most cited evaluation approaches in the literature. However even to refer to these logic models as ‘evaluation’ tools would be inappropriate since they were initially conceived as planning tools (Rogers & Fraser, 2014, p.3). That is – they outline a logical relationship between activities, outputs, purpose and goal (Rogers & Fraser, 2014, p.4). In this way, they mirror much of the thinking in the traditional PM paradigm where objectives clearly led to expected results (Crawford & Pollack, 2004, p. 649).

Indeed logic models often serve an accountability-for-results mentality, i.e. a question of ‘how’s my money being used?’ instead of emphasizing managing-for-results, which
asks ‘is my money having impact?’ (Ika & Lyvytnov, 2012, p.87). In effect by pursuing logic models, the wrong kinds of questions are often posed and the efforts of evaluators fail to capture whether a project is fulfilling its development goal (Walker, 2015, p.77). For this reason logic models have a way of quantifying program success in a linear fashion while neglecting critical process issues that could better shed light on the performance and improvement opportunities of a program. Even so, logic models like LFA have enjoyed permanence in the sector that has less to do with its impact and appropriateness and more to do with pandering to the preferences of donors (Biggs and Smith, 2003; Smith, 2000 cited in Golini et al, 2014, p.652).

However the models under this traditional evaluation paradigm have not come without their resistance. People-centered development advocates have gained considerable visibility in underscoring the importance of having more qualitative, process outputs (Mog, 2006 p.543). Indeed the 2000’s ushered in a paradigm shift from traditional ‘performance measurement’ evaluation towards outcomes evaluation (synonymous with the references to outcome measurement, impact evaluation or the impact movement). Under the outcomes evaluation paradigm, it is argued that there is a causal relationship worth exploring (Stern et al, 2012, p.6) by observing the more qualitative elements of a program.

The shift towards the outcomes evaluation paradigm also represented a move from efficiency towards effectiveness (Liket et al, 2014, p.172). This is a relevant point to the investigation as it signals a shift away from the traditional evaluation era fixated on outputs and efficiency. Rather, effectiveness advances the view of program evaluation as it considers whether the program is ‘doing the right thing’ (Bryce & Crawford, 2002, p.366). In other words, are programs fulfilling the development objective they had set out (Walker, 2015, p.77)? This is a key shift from PM thinking because it recognizes that non-profits and the ID sector are inherently social in nature and thus cannot simply apply PM’s iron triangle of measuring efficiency (Hermano et al, 2013, p.24). In its place, effectiveness endorses sustainability where economic, social and institutional goals, although amorphous, are of great concern (Angelsen et al. 1994; Bezuneh et al. 1995; Flora 2001 cited in Mog, 2006, p.532). In this way, the outcomes evaluation movement is more appropriate as it takes a step closer to recognizing the ‘soft’ nature of projects.

Moreover, instead of disappearing, efficiency remains very much enmeshed in the concept effectiveness (Bryce & Crawford, 2002, p.367). In fact, the conversation expands its view on outcomes as an immediate achievement of program goals or contribution to desired societal change (Campbell, 2002; Poole et al., 2001 cited in Liket et al, 2014, p.174) to also consider outcomes in terms of their short, medium and long term manifestations (McLaughlin & Jordan, 2004 cited in Carman, 2010, p.260). In effect, outcomes evaluation pays attention to the specific context of ID in a way that the former traditional evaluation era never did (Buckmaster, 1999, p.190).
The other noteworthy feature of the outcome evaluation paradigm that parallels the prioritization of effectiveness is a desire for learning, often referred to as ‘formative evaluation’ in industry terms (Chen, 1996, p.123). In effect, the outcomes evaluation movement advances the notion of program evaluation so that it is not merely about capturing results but also learning for the sake of improving programs (Ebrahim, 2003, 2005 cited in Benjamin, 2012, p.1227), and genuinely improving the odds of reaching the development goal. Certainly, this element of learning is pivotal as it “enables [organizations] to effectively respond and adapt in complex and rapidly changing contexts, such as those in developing countries” (Behrens and Kelly, 2008; Hay, 2010; Pearson, 2011 cited in Lennie, 2015, p.326). In others words, there are feedback systems and learning processes put in place with improvement as a goal (Buckmaster, 1999, p.192). This is a significant departure from the pre-dominant reporting function that evaluations were seen to have in the previous paradigm.

Although there is another paradigm shift left to discuss, it is widely cited in the literature that most non-profits adhere to outcome evaluation practices (Thayer and Fine, 2001, p.104). One of the reasons the evaluation paradigm remains locked in outcomes is due to the fact that most evaluation practices are still inspired to fulfill the interest of donors (Carman et al, 2008, p.53). Although each donor has varying evaluation preferences for evaluation conducted by non-profits, almost all donors share a transversal need to demonstrate results (Golini & Landoni, 2010, p.125). As some of the key promoters of the accountability movement as a means to demonstrate transparent use of funds, donors (often times governments, multilateral organizations etc.) are equally pressured to exhibit to the public that their monies are having impact (Chaves, Stephens, & Galaskiewicz, 2004; Froelich, 1999; O’Regan & Oster, 2002; Peterson, 1986 cited in Lee & Nowell, 2015, p.311; Buckmaster, 1999, p.188; Lennie et al, 2015, p.330).

Unique from the other two paradigms, participatory evaluation emerged from the development evaluation field to problematize the way evaluations are conducted under the current outcomes evaluation paradigm. Wallace puts forth a popularized notion of participatory evaluation where there is “shared control of procedural decision-making and stakeholder participation in all aspects of evaluation except for data analysis” (Wallace, 2008, p.202). The most salient feature of this paradigm is a call for more equitable participation of key project stakeholders in the evaluation process (Burke, 1998; Springett, 2001 cited in Suárez-Herrera et al, 2009, p.327). Instead of being donor-defined, program beneficiaries (communities and local organizations) are invited to take a larger role in the design and implementation of evaluation practices (Bamberger, 2000, p.98). This mirrors much of the sentiments of the critical paradigm that advocates for radical re-imagining of how ID is structured (Ika and Hodgson, 2014, p. 1186).

The key merit that this is cited to produce is increasing the co-ownership of the project while also ensuring that the program becomes more sustainable and thereby more
effective (Suárez- Herrera et al, 2009, p.324). Even so, the adoption of participatory evaluation has not been widespread in large part because its objectives are romanticized in theory yet difficult to achieve in practice (Rebien, 1996, p.153). The most notable obstacle has been the juggling of competing interests of stakeholders. Particularly, in an environment where non-profits are still pressured to fulfill donor wishes first, and beneficiary communities second (Ika & Saint-Macary, 2012, 431).

In this backdrop, it is challenging to pinpoint one paradigm that current program evaluation practice currently adheres to. On the one hand, the thinking at least in discourse has embraced more process oriented and participatory perspectives of the last two paradigms. However in practice, the frameworks are locked within the same outdated reporting-centered thinking of the traditional evaluation era. From this point of view, the true intention of program evaluation to capture performance and improvement opportunities of complex ID programs is at odds with the inflexible frameworks used to manage it. Put otherwise, the tools promote a notion of evaluation as rationally feeding into programs (Rogers & Fraser, 2015, p.2); not unlike the blueprint approach era of PM for ID. Meanwhile, the thinking that dominates the literature and empirical studies suggests that the reason why these frameworks have maintained permanence is to adhere to donor preferences (Liket et al, 2014 p.173). Nonetheless, this has not stopped discourse from promoting more innovative program evaluation approaches, even if program evaluation practice remains somewhat caught between paradigms.

3.8 PMD Pro: A Critical Analysis of PM4NGO’s Contribution to Program Evaluation

In this section, the study critically reviews the most concrete contribution PM has articulated towards ID programs: the creation of PM4NGOs and its guide PMD Pro. Indeed, when PM4NGOs was created, it was done with the intention of being able to bring PM knowledge to NGOs and INGOS in the development sector. PMD Pro was thus envisioned as a guide to bolster the PM capacity of development professionals (Culligan et al, 2011, p.1).

Within the guide, and of most interest to this study, there is a section uniquely addressing the ‘Project Monitoring, Evaluation and Control’ phase of a project/program. The phase is described to:

[Extend] through the entire life of the project and continually measures the project’s progress and identifies appropriate corrective actions in situations where the project’s performance deviates significantly from the plan (Culligan et al, 2011, p.17).

By this definition, the vision of evaluation communicates both a consideration for reporting on progress (akin to monitoring) as well as learning for program improvement purposes. The element of ‘control’ discussed in this chapter refers to responding to
change management, a feature that is not as evidenced in the Development Evaluation literature.

Although evaluation is recognized as a phase, the PMD Pro considers “the Monitoring, Evaluation and Control Phase as a background that extends from the earliest tasks of Project Identification and Design, all the way through the last tasks of the End of Project Transition Phase” (Culligan et al, 2011, p.56). More to the point, the guide underscores that evaluation questions should be interlaced within the project plan. Among the slew of questions the guide proposes are: “Who is responsible for collecting data, processing monitoring data, analyzing data, documenting results and communicating messages?” and “What resources will be required to complete the evaluation?” (Culligan et al, 2011, p.46). Seemingly, there is a tendency to deal more with operational matters. Based on this alone, the questions are more inclined to get at the efficiency of a program rather than the effectiveness.

This being said, one reassuring sign is the recognition that initial planning documents that feed into the ‘Monitoring Evaluation and Control Phase’ are considered ‘living documents’. That is – they are not regarded as static but rather embrace changes as they come (Culligan et al, 2011, p.48). Furthermore, the guide explicitly states: “organizations should choose their evaluation approach based on their learning objectives” (Culligan et al, 2011, p.61). This openness towards adaptive management and learning is significant as this kind of logic is not well established in traditional PM thinking. Indeed upon reviewing the Project Management Body of Knowledge (PM BoK) produced by the Project Management Institute, it also became evident that no such chapter on evaluation exists (PMI, 2013). This begs the question of where knowledge on program evaluation is coming from to input into PMD Pro?

The only practical framework that is concretely cited in the guide is that of the LFA, which we know already from the literature review to be a catchall planning, monitoring and evaluation tool created by donors in the ID field (Corti & Landoni, 2011 p. 50,); it is described as:

[The] first step in developing the full monitoring and evaluation plan for the project. The indicators and means of verification that are included for the logical framework will ultimately become the building blocks for the full monitoring and evaluation plan of the project (Culligan et al, 2011, p.58).

Beyond championing the exclusive use of the LFA, the guide only asserts three evaluation ‘approaches’ that are used extensively in the sector: the final evaluation, mid-term evaluation, and ex-post evaluation. The first is described to gage immediate outcomes the second to gage progress and offer feedback for corrective action and the last to observe impact following a period after program closure (Culligan et al, 2011, p.61).
3.9 Bridging the literature

Thus far the narrative should demonstrate some of the intersecting lines between PM, PM in ID and Program Evaluation. Upon reviewing the microevolutions in PM in ID as well as in program evaluation, it becomes evident their paradigm shifts mirrored one another. In other words where there is a symbiotic relationship between PM in ID, illustrated by Ika and Hodgson (2014, pp. 1183-1189), and the three evaluation paradigms found in Buckmaster (1999), Rebien (1999) and Suarez-Herrera, et al. The responsiveness of these fields to one another serve as reminder of why this study pursues a subjectivist lens. It can be argued that the way PM in the ID sector evolved clearly shaped the thinking and practice behind program evaluation.

Indeed, it is no coincidence that the traditional blueprint era of logical/rational approaches in PM for ID was mirrored by logic models in program evaluation. Likewise when the ill fit of the traditional era became evident, the switch to a more people-centered approach spurred a similar change in program evaluation. This was represented by a shift towards outcomes, which by definition considers the more intangible social benefits of programs. Lastly, the critical paradigm of PM in ID is very much emblematic of where participatory evaluation seeks to go. That is to say both of them advocate for questioning underlying assumptions in development. Both bringing in existential questions for consideration and yet neither having been adopted widely into practice. Figure 3 below, summarizes how the evolution of PM can be compared to the evolution of program evaluation. This includes characteristics from PM in ID that have clearly shaped the micro-evolutions within program evaluation.

However what the literature has also indicated is that PM practices that are used in the ID sector today were developed organically from within the sector. The lack of commentary from PM on how it can contribute and tailor itself to ID is evident in the literature, especially so on program evaluation. What can be abstracted from the literature, however, is that PM’s iron triangle of cost, quality and time is the traditional conception of measuring performance and evaluating for the success of a project. Much of what the iron triangle promotes is exactly what the conventional era of ID in PM and the traditional era of program evaluation encapsulate. However these are conceived as outdated paradigms in the literature. Instead more emphasis is placed today on the outcomes evaluation paradigm and people-centered paradigm. Within that space, PM4NGO’s guide ‘PMD Pro’ may one of the only tailored contributions to have come directly from PM. What remain to be answered is if this guide is in tune with those central paradigms of today and if in fact its contribution is meaningful. As such, assessing PM’s contribution against organic developments of PM in ID should feed into the first research question: to what extent has PM contributed to program evaluation?
In sequence, what is happening currently in the sector in terms of the outcomes evaluation, and even the participatory evaluation, paradigms can provide recommendations for PMD Pro. As such, observing these relationships should answer the second research question of: how can current program evaluation practices inform PM4NGOs?

The relationships described above can be seen in figure 3. The diagram shows one the one hand, the interaction between PM in ID influencing program evaluation practices. The most influential characteristics of each PM in ID era have been highlighted as shaping the different program evaluation paradigms. On the other side, PM is seen, where thus far the only concrete contributions towards program evaluation have been the iron triangle and recently the creation of PM4NGOs and its accompanying guide: PMD Pro. The dashed line around PM4NGOs represent the lack of analysis that has been carried out in current literature to determine its relevance to current paradigms. Differently, the dashed lines around outcome evaluation and participatory evaluation towards PM, show that it remains to be discovered whether they can contribute to PM.

Figure 2: Key Characteristics of Paradigm Shifts Within Project Management in the Development Sector (Ika & Hodgson, 2014, pp. 1183-1189; Buckmaster, 1999; Rebien, 1999; Suarez-Herrera, et al., 2009)
Chapter 4. Data Presentation

4.1 Semi-Structured Interviews

The interviews completed shed light on some of the particularities of program evaluation. They further highlighted some of the nuances and trends emerging in the field. Below is a presentation of the main findings, presented in sequence of the interview questions asked.

4.1.1 INGO Interview Findings

1. What evaluation approaches and methods are most commonly used in the organization?

A: “We’ve used mixed-method approaches because you can triangulate and cross-validate both qualitative and quantitative data. We also use some case study methods like most significant change”.
   • War Child Officer

“Most often there is a preference for mixed methods [...]. This is usually done when the donor is not keen on the necessity of evaluation. Impact evaluation should be the preferred tool and method used but mix methods are usually the ones that get used.”
   • Plan UK Officer

It was important for this study to understand how the use of tools and methods influence the practice of program evaluation in development. A common response to which methodologies were applied was mixed-method approaches. They are said to collect a broad range of data; balancing the desire for hard results and in-depth investigation. One unexpected response justified the use of mixed-methods to collect and process data because it was the approach that took the least amount of effort and money.

This question had a follow-up / sub-question: If possible, could you give examples of tools used by the organization to evaluate?

A: “We started to talk about using more theory of change models and qualitative methods. So we’ve opened up to being more innovative”.
   • Restless Development Officer

“Using purely qualitative [methods used] is quite rare. There is an assumption that if you have a large or multi-country project you should at least have some [qualitative methods].”
   • Plan UK Officer
Respondents agreed that the types of tools utilized during an evaluation depend on the type of program that is being carried out. Most organizations admitted to using mixed-method approaches. They strived to balance quantitative and qualitative tools, although there is a growing emphasis for the qualitative. Differently, at least two of the organizations referred to specific evaluation methods that they more consistently adhered to. One example of this can be highlighted by the use of process tracing as a tool to measure advocacy work impact (Restless Development Officer, 2015, personal interview). It also appears that the scope of the program influences the use of tools. One respondent described how programs with larger scope necessitate more tools to pull off the reporting and evaluation functions.

There was also talk of adhering internally to evaluation frameworks beyond the LFA. Among them, Theory of Change (ToC) and to a lesser degree Most Significant Change were often mentioned.

In reflecting about approaches to evaluations, organizations also commonly referred to the importance of integrating evaluation thinking from the planning stage. In this way, they can end up with a more effective evaluation. As such, the methods and tools selected should also be able to respond to the initial baseline set out in the program.

2. **Are the evaluation approaches and methods you use introduced by you, the organization, or by donors?**

A: “There is no particular evaluation approach suggested by donors. Normally if you go to the donor and say you are using mixed method approach they will be fine with it.”

  • War Child Officer

“Donors are relatively flexible, and can encourage the implementing agency to make choices based on what’s best for them. Donors don’t tend to get involved in detail. They are not prescriptive of the methods. More they are interested that it happens and that [the evaluators] are external. [European Commission] prefers external evaluations, for example”

  • Plan UK Officer

“There are not really a lot of scenarios where donors are asking for a particular method or trying to impose standards. More than often there is a steer from donors to the evaluation questions they want answered.”

  • Action Aid Officer
“We work a lot with corporate donors. They have different priorities, aside from measuring impact [...]”

- Youth Business International Officer

The range of responses to this question brought to light a few nuanced narratives. Still, one point did reoccur: donors cannot be considered a homogeneous group. That is to say their preferences when it comes to how program evaluations should be conducted vary broadly. Between governmental, multilateral and corporate donors there is a wide-ranging spectrum of evaluation approaches from more hands on to more lax. Although it would be ill advised to draw a generalization, all respondents concurred that donors were not stringent about prescribing methods used to collect data. What was less clear to understand was whether INGOs and donors shared a vision of the driving ethos behind program evaluation?

While the practicalities of how evaluations are conducted are suggested to be of little interest to donors, what did seem to hold importance is that INGOs comply with reporting frameworks such as the LFA. The LFA in particular was cited in several of the interviews.

Another key point that was mentioned with regards to governmental donors is the fact that many of them are in the practice of commissioning external evaluation (also referred to independent evaluations); seemingly a growing trend in the sector (Snyder and Doan, 1995, p.150). All of the organizations interviewed state that donors have increasingly expressed a preference for evaluation of programs to be carried out by a third party, i.e. external consultant. Some donors go as far as having this etched into the guidelines and agreements when grants are awarded.

In the case of organizations that work with corporate or private donors, the prescription of tools and methods for monitoring and evaluation seems to be almost non-existent. One respondent explained this as corporate donors having different motivations for financing ID programs (mostly tax-breaks). Moreover, the call for accountability in the use of funds seems to be less rigorous. The rationale for undertaking an evaluation is thus more centered on telling a story about what is being done. In short, where demonstrating hard results contributes to bolstering the positive corporate image of the financing company.

3. **How would you describe the organization’s approach to evaluation?**

A: “The organization has a global level commitment to evaluation.”

- Restless Development Officer
“It’s a matter of compliance. You are committed to do it and you should do it, even if you do not generate learning. In other cases it is an area where the organization wants to dig deeper.”
  • Plan UK Officer

All respondents were able to recognize the invaluable role of evaluation to the organization and improving the quality of its programs. In delving deeper into the whys of conducting evaluations, a common range of answers surfaced. Among the more recurring buzzwords include: accountability, learning and improving program and organizational effectiveness. However some admitted that the particular focus will depend on the nature of the program, the type of funding (restricted or unrestricted) and the donor behind that funding.

a. What is the organization's purpose in carrying out an evaluation and what use(s) does it serve?

A: “We say participation is important. Young people need to be a big part of our evaluation strategy. We want to be able to move away from staff just doing evaluation. How can we have young people be a part of that? […] if you give them these clunky confusing tools it becomes hard for them to actually produce meaningful results. It’s a paradox”.
  • Restless Development Officer

“The classic function of accountability and learning. But this accountability of demonstrating to donors is different dependent on the donor. Other times it is for showing impact.”
  • Youth Business International Officer

Most organizations underlined the importance of going beyond treating evaluation as a simple donor compliance measure for accountability. Indeed some organizations saw that in addition to accountability, it was deemed supremely important for the organization to internally define a more provocative and innovative ethos to evaluation. That is to say that some of these organizations seek to move evaluation practices forward. They look to lead by example by introducing more participatory techniques to engage beneficiaries and show impact. Some organizations have gone as far as having their own evaluation policy, an emerging trend in the sector.

4. Where does program evaluation begin and end?

A: “We start with the design of the program, defining evaluation questions at the beginning, considering what are the questions we want to answer.”
  • War Child Officer
“Start as early as you can. Some people talk about monitoring and evaluation. But it’s inherently linked into the planning. What you are evaluating is linked to your intended results. Objectives that you set up right from the start.”

- Restless Development Officer

“[…] large programs that run across several years and are very strategic, start from the beginning of the project and end after the end of the project.”

- Plan UK Officer

For most organizations evaluation is invariably tied to the program planning phase. The goals and outcomes of an evaluation will respond to what was done at the front-end of the program since the goal is to start as early as possible, beginning with baseline measurements. One organization also underscored that the mid-term evaluations were quite significant as they give program managers the opportunity to gage progress and make appropriate adjustments to steer the program towards success. In discussing the end of an evaluation, there was a variation of responses. Most hinted that evaluation budgets and donor driven timelines often pushed end evaluations to be conducted upon closure of the program. In effect, the opportunities of gaging real impact (observing mid to long term outcomes) via an ex-post evaluation, i.e an evaluation conducted some time after the close of a program, are lost.

5. In conducting a program evaluation, what are the criteria normally guiding the evaluation (key words)?

A: “Generally we use the DAC evaluation criteria”

- War Child Officer

“Efficiency, effectiveness, impact, relevance and sustainability those are the 5 evaluation criteria by the OEDC. Internally that is in our guidelines. ToRs tend to be based around those.”

- Restless Development Officer

“OECD - DAC criteria: [We don’t] expect an evaluation to address all the criteria. Even a final evaluation cannot address them all. The idea is to focus on specific criteria i.e. Impact, or sustainability, or relevance. The approaches are very different and will depend on the project itself.”

- Action Aid Officer

“For us evaluations help build a body of evidence, particularly in thematic areas: education, girl empowerment; as well as for programmatic evidence for areas we are interested in working in the future.”

- Plan UK Officer
When conducting the evaluation of their programs, the majority of respondents cited the criteria set up by the Development Assistance Committee (DAC) of the Organization for Economic Co-operation and Development (OECD). This makes particular reference to five criteria: effectiveness, efficiency, sustainability, impact and relevance. Beyond this, there are also internally defined criteria that are more often than not aligned with the organization’s strategic and thematic areas of interest. Some organizations even went as far establishing global criteria to more easily run a meta-evaluation of all their programs. By focusing on these criteria, organizations are better able to plan for the future and observe what areas need more attention.

6. What are the strengths of program evaluation in the sector today and areas in which needs to improve?

A: “The sector in itself focuses too much on reporting and accountability, and the learning questions, which is what [we value] in evaluations, aren’t always as clear as they need to be.”
   • Action Aid Officer

“Balance should be tipped more in favor of monitoring and evaluation”
   • Plan UK Officer

“There are more people that are focusing on M&E, more investment which is good”
   • Restless Development Officer

“[…] there is a lot of movement in terms of […] communities of practice. This is really good, the debates on how to carry out evaluations”
   • War Child Officer

The interviews showed a mixed audience with respect to whether evaluations are too heavily focused on accountability and reporting instead of lessons learned. Some articulated that there should be more of a tipping point towards more genuine monitoring and evaluation practices instead of simply reporting about the implementation and execution of programs. The advantage that was echoed across the board is the tremendous momentum around program evaluation. All respondents perceived this movement as a step in the right direction.

Upon finishing interviews, interview transcripts were scanned with a key word search to assess the frequency of certain key evaluation buzzwords. To visually show this, a word cloud has been created; where the frequency of the term is represented by their size. The more a word was said in an interview, the larger it appears in the word cloud.
As evidenced by the word cloud, the terms most referenced in the interviews are donors, followed by: impact, learning, framework and accountability. The takeaway from this is that the relationship with donors is influential when it comes to the program evaluation practices of INGOs and driving trends in the sector. What remains more questionable is whether donors are central to the conversation because they co-opt the formation of program evaluation practices (Snyder and Doan, 1995, p.149) or because they are simply one among many relevant stakeholders in program evaluation.

As far as learning and impact go, these words reflect the desire for INGOs to make evaluations more meaningful; in tune with thinking of the outcomes evaluation paradigm.

4.1.2 Key Informant Interview Findings: Ruggero Golini

1. What does PM intend to offer the ID sector?

A: “PM has been quite detached from the development sector since the beginning. Because of this the development sector created its own PM approach. There have been approaches developed ID that are not known in the PM practice. So the two streams they have developed independently. The kind of contribution that PM could bring to ID is not 100% clear. The more you think of [traditional] PM the more difficult it becomes to export the techniques. This is because ID projects are soft, so they are not necessarily concerned with controlling the project with [traditional PM tools]. They don’t even have the data, it’s really too much. I don’t think the ID sector is aware of, or really wants to be influenced by PM work.” (Golini, 2015)

Golini poses the idea that there is a clear separation between the development of PM and ID in the treatment of evaluation. He also reinforces the fact that the needs of
programs stemming from the ID sector differentiate to the traditional PM-serving sectors.

2. **To what extent do you think PM’s tools have been adopted by INGOs?**

A: “It is hard to say as there is little info. In the analysis we carried out, PCM was adopted by more than 80% of the organizations polled, whilst PMDPRO and PMBoK were adopted 40%. Budget is a big problem because certification is expensive.

Golini’s comments recall that there is a limited range of tools and methodologies from PM, such as the PCM, that have been adopted within the ID sector. Even so this is but one example, most traditional tools have not experienced traction in the ID sector. Moreover, certain traditional PM tools and methodologies remain inaccessible due to budgetary constraints.

3. **Do you think PM is a field, a sector a discipline or merely a set of tools?**

A: “It’s a tough question. I’m always fascinated by the fact that, at least at the beginning, PM attempts to make a process of something that is not a process.” (Golini, 2015)

“PM cannot be just described as a process, but it’s trying as much as possible to standardize and to create processes out of something that is difficult to standardize. The tools are just a way to implement or support this process. If you look at the PM BoK it is basically just processes. Just referring to the tools is too restricted. I think it’s more of a methodology, or at least that is very the PM BoK way to interpret PM.” (Golini, 2015)

“The processes approach applied to NGOs can be risky because it creates a very heavy structure. Unless you are a very big NGO that has the skills and budget to sustain the overhead, a competence based approach could be more suitable.” (Golini, 2015)

The answer leads towards a definition of PM, but still leaves it open to interpretation. One of the strongest points made was the nature of PM as a process more than a set of tools, a methodology or a skillset. This capture a more fluid and dynamic perception of PM that could fit ID better. The other important takeaway is how this affects the application of PM in ID. Golini emphasized that heavy bureaucratic structures can be cumbersome and negatively impact management practices of INGOs. This is particularly pertinent for INGOs as they are the largest and most significant implementing agents of ID programs. As such, they need to rely one way or another on a suite of management tools and functions in order to operate.
4. Program evaluation is the area of particular interest in our research; to what extent has PM been able to contribute knowledge in this area?

A: “There is a big difference between programs carried out by NGOs and traditional programs.” (Golini, 2015)

“My sense is that PM provided some tools to keep the project under control, to help frame some accountability standard. NGOs are subject to accountability from the donors who want reports in a certain way with specific indicators.” (Golini, 2015)

Although the insight was limited, Golini was able to highlight some niche areas where PM could contribute. For example, PM tools may help donors monitor their ‘investments’ and ask for better standards of accountability. In his view, this could be one of the most concrete and useful contributions PM can make to ID.

5. Do you think there is a space for PM to learn from ID?

A: “I think there is growing interest from the PM community in the ID sector. More of the challenges emerging in the development sector (many stakeholders, many soft features) are starting to emerge more in the traditional PM community. I think the PM community is a quite generous community so they also want to contribute to society, to social good. There is a willingness to support social impact. So I think there is knowledge that can flow vice versa. Both fields can benefit from staying in touch, without one dominating the other.” (Golini, 2015)

What seemed to be a common thread in the interview, there is a learning space where PM and ID intersect. Although previous answers show that there is a variation between the desired outcomes in PM influenced industries and the ID sector, it doesn’t mean that both disciplines cannot learn from one another.

6. How is evaluation perceived in traditional PM?

A: “As far as I’ve seen there are many situations in the private sector where more [traditional] PM [tools are] applied. There are companies that work by projects – the evaluation of the projects is directly connected to their accounting systems. For sure there is an economic evaluation that is very precise. You need to track all the costs of the projects. My sense is that NGOs don’t have a strong project focus even though they are run by projects. “ (Golini, 2015)

“NGOs seem not to have a traditional project focus and they do not do a front end evaluation (cost of the project) - and it is harder for them “ (Golini, 2015)

Golini claims that there is a difference in the ultimate goal of evaluation in both disciplines. He characterizes INGOs as having a lack of precise economic
evaluations. This is reasoned by the tendency for their objectives to be more socially inclined. In short, INGOs in theory place a lot of emphasis on seeing social ‘soft’ changes take place among beneficiaries. The economic outcomes they need to serve deal more with responding to funders.

4.2 Content Analysis

4.2.1 Content Analysis Findings

To conduct the content analysis, a number of keyword searches were run through the ToRs. Keywords were selected on the basis of a technique called ‘latent coding’, which involves the researchers inducing the different themes from the texts (Ryan & Bernard, 2003, p. 86). Considering the popularity in the program evaluation discourse terms were induced from the literature review at first. Preliminary observations and first read-through of the ToRs also uncovered word repetitions that were pertinent to the study. The discovery of word repetition is considered an empirical technique that draws out the more salient terminology in the content that is being analyzed. The searches were namely divided into two cluster areas: DAC OECD criteria and evaluation approaches.

The first refers to evaluation criteria produced by the DAC OECD including: effectiveness, efficiency, sustainability, relevance and impact (OECD.org, 2015). The criteria first came to our attention in the process of conducting interviews, and had previously been not been noticeable from the literature review. They refer to the five points settled under the 2005 Paris agreement of the OECD to establish more far-reaching and monitor-able actions in the way that development aid and cooperation are managed and delivered (OECD.org, 2015).

ToRs were deemed to adhere to DAC criteria if any of the five criteria were listed in the evaluation purpose of the ToR or otherwise stated outright in the research questions portion of the document. Given that each organization had a unique way of structuring the ToR, the way the DAC criteria are presented looks different in each ToR. This being said, it’s important to note that as ever-present as the DAC are in calls for evaluation, not all five criteria are always present. Of the ToRs analyzed, 80% featured some degree of DAC criteria. In this 80% where criteria were present, a combined 62.5% of ToRs name at least four if not five criteria. Thereafter, 25% of ToRs list only one of the criteria, and far fewer ToRs list two or three. This can be seen in figure 1.
Seeing the majority of ToRs adopting at least four to five of the criteria, we then computed which of the criteria emerged as most cited in the evaluation aims of the ToR. The results seem to suggest a relatively equal favoring of the criteria with effectiveness only slightly edging out over the others. This signals that INGO’s calls for evaluations fulfill a diverse set of program intentions. Figure 2 shows these percentages.

The word search for criteria driving evaluations was run again only this time the goal was to count the frequency of the terms throughout the ToR. This search was expanded to include the same DAC criteria of earlier as well as adding the extra key word of learning, which was one of the salient terms noticed during the application of the word repetition technique.
Once criteria were counted, ranked and weighted, it emerged that effectiveness features most prominently at 29%. That is to say that calls for evaluation emphasize effectiveness repeatedly, followed by the criteria of learning at 20%. The presence of both these criteria, but especially the latter, signal that evaluations are interested in more than just demonstrating accountability.

Most of the other DAC criteria follow closely behind but the one criteria noticeably lagging is that of efficiency, with a frequency of 8%, the implication here being that evaluations do not consider efficiency as a major driver. Figure 3 represents graphically the distribution of the ranking.

Figure 6: Ranking of Criteria in ToRs

The second cluster area was categorized as approaches to evaluation. The keywords complacent to this category were also derived from hints of popularized terms in the literature and observed tendencies during preliminary ToR review. These included words such as: theory of change, qualitative, quantitative, participatory, results based, logic model, mixed methods and formative. The reoccurring nature of these terms made it important for the researchers to observe. Moreover each of these terms has been commonly cited in the evaluation literature and are equally indicative of the paradigm shifts lived in the sector. What the findings revealed is that most organizations emphasize use of qualitative data collection processes and the adoption of Theory of Change (ToC) as a framework to capture change caused by the program. Interestingly enough, quantitative measures were only emphasized half as much as qualitative ones.

Lastly, a few additional keywords were searched independent of any category to gage their importance in evaluation calls. Jargon such as evidence, recommendations, outcome and outputs were among the terms included in this search. For example, 70% of ToRs made reference to outcomes versus outputs appearing in only 45% of
ToRs (Appendix 6). If this is taken as an indicative trend in the sector, then it is concurrent with the idea that evaluation is currently within the outcomes evaluation paradigm.

The high instances of terms ‘evidence’ and ‘recommendations’, appearing in 85% and 70% of ToRs (Appendix 6) respectively suggest that calls for evaluation are encouraging more feedback, not only for internal improvement purpose but also for external reasons. The term ‘formative’ evaluation is often used to imply feedback driven evaluations (Marshall and Suarez, 2014, p.1035) yet what the content analysis also revealed is that the language of evaluation is not always so widely adopted with the actual term ‘formative evaluation’ only appearing in 5% of the cases. So although the appearance of keywords certainly can tell a story about the discourse, the varied adoption of terms hints to a lack of standardized practice in program evaluation. Figure 4 highlights these findings.

![Figure 7: Approaches to Evaluation](image)
Chapter 5. Data Analysis

5.1 Integration of PM practices in Program Evaluation

The first research question that was posed for this study asks to what extent are PM evaluation practices integrated into program evaluation in the ID field? The proposition that the study has put forth, and that will be gaged here, is that the contribution has been insufficient. In answering this question, the study draws from its interviews with INGOs, interview with scholar Ruggero Golini and the critical literature review. The analysis intends to show that PM developed organically within the ID field, largely independent from the PM discipline. As a result, PM has not made any significant contributions to program evaluation, until the development of PM4NGOs. This can be gaged by the level of adoption of PM practices within the ID sector.

5.1.1 Creation of PM tools in ID: Parallel Growth

One of the key takeaways from the literature review and the interview with Professor Golini is that PM has been a fundamental vehicle through which to deliver donor sponsored ID programs, precisely because ID is a project-based field (Ika & Hodgson, 2014, p. 1182; Golini, 2015).

However the paradox is this: although ID is a project-based sector, its PM practices were organically derived from ID rather than adopted from the PM discipline. This proposition was first signaled by the overall scant treatment of PM in ID (Ika et al., 2012, p.105) but reaffirmed by assertions in the literature and Professor Golini: ID took it upon itself to develop a suite of PM practices that were more suited to the particular needs of the sector (Hermano et al., 2013, p.29). Aside from adopting a few tools such as the internal rate of return and cost benefit analysis, as seen in Table 1, the uptake from PM falls short when it comes to program evaluation. In light of PM’s inability to provide tailored tools to ID, the ID sector proceeded to create their own tools and approaches (Hermano, et al. 2013, p.23; Corti & Landoni, 2011, p. 46).

In fact the literature on PM’s contribution to program evaluation is virtually non-existent. Part of the reason behind this is because program evaluation has been an underdeveloped area within the imaginary of traditional PM. This became especially clear upon revising the PM BoK produced by the PMI where no such chapter on evaluation exists (PMI, 2013). Golini compliments this perspective by asserting that at best, PM offers a planning and reporting tool: the earned value system. However he also goes on to say this kind of tool is inappropriate for ID programs as it was developed with traditional ‘hard’ PM programs in mind. Meanwhile program evaluation needs PM tools that are more cognizant of the ‘soft’ nature of ID programs (Golini, 2015).
Instead what the limited literature does observe is how program evaluation tools have been produced by the ID sector itself. Although envisioned as catchall planning, monitoring and evaluation tools, as opposed to exclusively evaluation, tools like the Project Cycle Management (PCM), LFA and Results Based Management (RBM) are seen to “provide better management control through appropriate links to the ongoing operation of performing organizations provide more stability” (Corti & Landoni, 2011, p. 46). Yet, even though PM has done nothing to contribute directly, as a body of knowledge or set of tools, to program evaluation, indirectly these frameworks are inspired by the logical and instrumental rationality of PM (Ika & Hodgson, 2014, p.1187), as can be exemplified by the history of the LFA, whose origins stems from the US military a sector traditionally linked to the PM field (Bakewell & Garbutt, 2005, p.1).

5.1.2 Trying to bridge the gap – PMD Pro

The narrative until now has demonstrated that although ID has employed its own PM practices to carry out ID programs, these practices have exclusively been inspired by the ID sector and not by PM; significantly more so when it comes to program evaluation (Ika et al., 2010, p. 63). If we were to answer the first question on this information alone, the response would affirm the proposition that PM’s contribution to program evaluation is insufficient. However the story does not end there. Indeed come 2011, a new association was formed to address the needs of international and national NGOs as an offshoot of PMI: PM4NGOs.

What the establishment of PM4NGOs intended to do was create a concrete tailored offer from PM to ID that encompasses a level of professionalization that was missing in the field (Brière, et al., 2014, p.116). The professionalization sought by PM4NGOs is focused on delivering a specific certification that unites Project Managers from INGOs across the world (Hermano et al., 2013, p. 23), as well homogenize practices known to the field of ID (Marinho, 2015, questionnaire).

To gage the influence that the PMD Pro Guide and PM4NGOs has had on the way program evaluation is implemented, it is important to go back to what has been said in the literature review. From the analysis of the guide it can be extracted that little attention has been paid to evaluation, with much of the discussion championing iterative changes to a program as it progresses (i.e a function of monitoring) (Culligan et al., 2011, p. 36). In fact, the guide seems to downplay the importance of, what they refer to as, the ‘design, monitoring and evaluation phase’, claiming that it “has sometimes overshadowed the importance of other phases in the life of the project” (Culligan, et al., 2011, p. 16).

Certainly, the PMD Pro Project Phase Cycle seen in Figure 5 highlights the idea of evaluation permeating throughout the program planning and implementation. Yet
aside from positioning evaluation within the frame of a program, the guide does not provide a great many options for evaluation approaches. In fact the only tool that the guide details is the LFA (Culligan et al., 2011, p. 29; Hermano et al. 2013, p.27), a tool that, as previously discussed, has been criticized for being confusing and restricting in its approach (Couillard et al., 2009, p. 32), and less than helpful for certain types of evaluation, like impact/outcome evaluation approaches (Gasper, 1999, p. 2; Bakewell & Garbutt, 2005, p. 11).

The PMD Pro is also outdated in the way it emphasis the supremacy of achieving the program’s goals, outcomes and outputs while meeting the iron triangle of time, cost and quality (Culligan, et al., 2011, p. 9). In fact, the analysis carried out leads the researchers to believe that the guide constructs an image of success that is more about outputs rather than outcomes. This adherence to the iron triangle echoes much of the ethos of traditional PM as opposed to more contingent approaches of PM in ID where the outputs and outcomes are not measured in terms of time, cost, quality, but more in terms of the impact had on the beneficiaries and the construction of solutions based on participatory approaches (Ika & Hodgson, 2014, p.1888).

Where the guide does have redeeming qualities is within its stress placed on iterative learning (Culligan et al., 2011, p. 36). Although this kind of learning is conceived more as a way to keep programs on tracks, it does pander more to the visions of an outcomes evaluation paradigm. This being said, there is still room for the guide to provide more concrete guidance on the kinds of learning questions that would allow INGOS to gain more meaningful insight into what “constitutes effective, useful, and professional evaluation practice” (Lennie et al, 2015, p.327).

By this measure, there have not been any great leaps between what the traditional PM body of knowledge offers and what PMD Pro is proposing. One possible explanation
may rest with the fact that PM4NGOs reiterates much of what is stated in the traditional PM knowledge: “including the philosophy, tools and techniques that have been developed in the PM discipline over the last 20 years” (Hermano et al., 2013, p.27).

Consequently, the added value of PMD Pro is undermined in light of its alignment with the conventional PM era of ID that stresses for “a well-structured pattern of efficient techniques that allow people to work together while bringing into focus the project’s objectives” (Corti & Landoni, 2011, p. 45). This is directly at odds with the more broadly adopted people-centered approach to PM in ID (Ika & Hodgson, 2014, p.1188).

Perhaps most problematic however, is that the guide is articulated as focusing on a project-level approach as opposed to program level (Marinho, 2015, Questionnaire). This delimits the value of the guide quite significantly since the majority of development initiatives are communicated through programs rather than projects as part of the sector’s commitment to bring about more holistic change (Binnendijk, 1989, p.207). Seeing this, PM4NGOs is in the process of putting together a new guide ‘Program Management in Development Guide’ to be released in 2016 (Marinho, 2015, Questionnaire).

Beyond the guide’s integrity and added value, the real impact of the guide is also at question since little research exists to attest to the guide’s adoption among the INGO community. In fact, of the six INGOs interviewed for this study, only two had heard of PM4NGOs and the PMD Pro guide, yet both had no knowledge of what exactly it entailed (Interviewee 4, 2015).

Although the association asserts that it is collecting data at present to evaluate the effectiveness of the PMD Pro program, no concrete evidence exists to date (Marinho, 2015, Questionnaire). Nonetheless, Figure 6, pulled from the organization’s website, did provide some indication of PM4NGOs reach. According to their figures, 10,936 professionals in the ID field have been certified with the PMD Pro in over 100 countries (Fig. 6), a relatively low number of adoptees, compared to similar associations in the PM field like PMI or APMI who have an estimate of 17,000 new people vying for certification each year (Project Management Institute, 2015, p. 4).
This being said, it may be too early to qualify PMD Pro since PM4NGOs only came into existence in recent years. Instead, the study recognizes an opportunity for PM4NGOs to more accurately identify and capture industry trends occurring in program evaluation. In the next section, much of the background and emerging trends of program evaluation in INGOs will be established. Ultimately, the set of recommendations at the end of the data analysis will make suggestions to how PM4NGOs can update its program evaluation offer.

5.3 Lessons Learned: Program Evaluation Practice

Having understood PM’s contribution to program evaluation practice as being outdated and stuck within a traditional vision of PM in ID, the study now proceeds to answer the second research question: **how can current program evaluation practices inform PM4NGOs?** To accomplish this, the following sections will first look to establish what program evaluation looks like today, what paradigm is program evaluation practice adhering to and ultimately what recommendations can be given to PM4NGOs.

5.3.1 Variation in Program Evaluation

As Henry and Mark (2003) put it: “we know remarkably little about how evaluation is being practiced, why it is being practiced, by whom and where it is being practiced, and to what effect” (cited in Carman, 2007, p.69). As a result, one of the most striking observations to come out of the study has been the variety of ways that program evaluation manifests across INGOs (Patton, 1986 cited in Rebien, 1997, p.440), and even within a single organization. What is less obvious is how much of program evaluation is standardized industry practice and how much of it is left to arbitrary interpretation and experimentation of INGOs.
What the interviews with INGOs revealed is that there is a combination of both. For one, the sophistication and professionalization of program evaluation is exceptionally varied (DiMaggio & Powell, 1983; Meyer & Rowan, 1977; Sahlin-Andersson & Engwall, 2002 cited in Marshall & Suarez, 2014, p.1036). All organizations interviewed possessed units dedicated to program evaluation, however each of them retaining their own unique features. The scope given to evaluation units alone hint at this, with most adopting variations of: Monitoring & Evaluation (M&E), Monitoring, Evaluation and Learning (MEL) or Monitoring, Evaluation, Accountability and Learning (MEAL) (Respondent 3, 2015; Respondent 2, 2015; Respondent 4, 2015). These units were at times seen as functioning independently, operating in tandem with program teams or in other cases organizational effectiveness teams. The degree of sophistication with program evaluation practices is also diverse, with some organizations going as far as having their own organizational policies on evaluation and even tailored methodologies (Respondent 5, 2015).

5.3.2 Donor Influences on Program Evaluation

The different manifestations of program evaluation could also be related to the funding source of an organization. Once more, although ‘donors’ as a term can be used to generically categorize those entities that fund non-profits in the execution of development programs, they are not a homogenous group (Lee & Nowell, 2014, p.310). Rather their unifying quality is that by holding the money, they also have the power to drive the ID sector in any direction (Golini & Landoni, 2014, p.124). In effect, donors can exercise the power they have to shape program evaluation practices (Bamberger, 2000, p.96). In spite of this, donors do differentiate in the unique interests, funding schemes, preferred frameworks and policies (David, 2006 cited in Carman, 2010, p. 259).

Most large INGOs, as concurred through our interviews, have a tendency to receive the bulk of their funding from a mix of large institutional donors, among which include: the UK’s Department of Foreign Affairs and International Development (DFID) and the European Commission (EU) to name a few (Respondent 5, 2015; Respondent 6, 2015). This kind of institutionalized funding can come with any range of conditions depending on the donor (Marshall & Suarez, 2014, p.1036). Nonetheless, that some may be more hands off than others does not negate the fact that all of them require INGOs to comply to some form of reporting framework, typically of the LFA or RBM variety (Rogers & Fraser, 2014, p.4). Yet, having to meet reporting requirements does not necessarily mean that donors impose methodologies when it comes to what approach non-profits should pursue to collect, compile and analyze data for an evaluation (Respondent 1, 2015; Respondent 2, 2015).
Yet some of the INGOs interviewed had a different funding basis and therefore a more nuanced perspective towards program evaluation. One case explained that a significant portion of their funding stemmed from corporate donors. Based on this organization’s experience, corporate funders were seen as having a much more superficial and simplified understanding of evaluation and as result, required far less rigorous reporting (Respondent 3, 2015). What is stressed instead is the primacy of being able to demonstrate results; presenting convincing numbers that would bolster the corporate’s public image (Lee & Nowell, 2014, p.311). In effect, for this INGO the evaluation approach for corporate donors was more traditional and ‘output’ focused than the kind of evaluation they would conduct for the institutional donors mentioned earlier.

Another INGO interviewed shared yet another experience entirely. They explained how their funding came pre-dominantly from a foundation that gave them the freedom to design an evaluation approach most suited to their needs and purpose as an organization. Consequently this INGO had developed a markedly participatory approach to evaluation that strongly advocated beneficiary involvement. Yet, their approach to evaluation took a turn when the economic crisis hit and the foundation was forced to significantly cut its funding. This left the INGO in a position needing to rely on a lot more institutional funding and in effect compromise some of the freedom they enjoyed previously. Although the organization was able to adapt to its new situation, there was recognition that aligning donor requirements and frameworks with the visions of INGOs can be challenging (Respondent 6, 2015).

The most fundamental point to draw from these different examples is that INGOs can be dealing with a host of donors, each of them possessing their own standards (David, 2006 cited in Carman, 2010, p. 259). Beyond simply responding to their needs and filling out their frameworks, the reality is that INGOs must put in a great deal of extra work simply to meet all their donor needs (Bamberger, 2000, p.97). In an ideal world evaluation requirements would be more streamlined so that the whole process could be a great deal less cumbersome for INGOs.

However since this is not the case, INGOs are in a position of chasing the money. As a result their program evaluation practices try to mirror and respond as best they can to donor requirements. Indeed, most INGOs interviewed cited that they have developed ‘global indicators’ in the last years to apply transversally across their programs (Respondent 1, 2015). In creating these indicators the organization is able to more clearly run a meta-analysis and see how programs are performing overall (Hanssen et al., 2008; Uusikyla and Virtanen, 2000 cited in Lennie et al, 2015, p.338). The trade-off, however, is that creating fixed global indicators might not respond to the every need of individual programs. If indicators are ill fitted to accurately capture what is going on – then what purpose does evaluation serve?
Indeed, if logic models continue to be what donors push out, then there is little room for more participatory evaluation approaches to incubate. Among the INGOs interviewed, most believe that donors have been earnest about pushing the discourse towards more participatory evaluations, however little else has been done to translate those good intentions into concrete action (Respondent 2, 2015). The reason behind this may be because participatory evaluations require that evaluators break away from the ill fitting indicators and frameworks embedded in most evaluations where beneficiary input was lacking (Lennie et al, 2015, p.332). Instead the proposition is that such indicators emerge organically from the communities where programs are targeted (Respondent 2, 2015). To pull this off, more time and money would be required, and not least alignment with baseline indicators set out during the planning stage of the program.

5.3.3 Commission External Evaluations

Despite their differences, there are some common trends that INGOs do share with regards to program evaluation practice. One of the most notable is the growing reliance on external evaluations; i.e commissioning program evaluation out to a third party.

This growing phenomenon of external evaluations comes in the wake of accountability calls and pressures on donors to demonstrate transparency (Snyder and Doan, 1995, 149). As Snyder and Doan would describe: donors often advocate external evaluations because they are “assumed to be more objective, less wedded to the project, and thus more able to identify its shortcomings” (1995, p.148). The mere existence of ToRs reaffirm that INGOs and donors are increasingly commissioning external evaluations. In these calls for evaluations, INGOs provide guidance and clear parameters on what exactly they are looking to evaluate. However, of the twenty ToRs revised, only five specified a particular methodology for external consultants to use; and even those five stemmed from the same organization; i.e an exception to the rule (Appendix 4). Rather what most organizations reinforced in the interviews, and what certain authors in the literature concur, is that donors are not stringent with INGOs about methods (Respondent 1, 2015; Respondent 2, 2015) and neither are most INGOs when they issue out ToRs for external evaluation.

This autonomy granted for methods is a dubious point as it begs the question of whether INGOs or donors commissioning out evaluations can verify the quality, rigorousness and legitimacy of the results produced by the third party (Naccarella et al., 2007 cited in Lennie, 2015, p.326). Moreover, it adds one more middlemen in the process of evaluating a program. Effectively, what is communicated and then interpreted is vitally important. Organizations that issue out ToRs for external evaluation with little instruction feed into the perception that evaluations are done poorly; giving disproportionate attention to performance rather than learning opportunities for program improvement (Christensen & Ebrahim, 2006; Cutt &
Although all organizations interviewed use some combination of in-house and external evaluations, the growing instance of the latter is significant. It ratifies the need for program evaluation practice to have some structure and continuity.

5.3.4 Adopting the DAC Criteria

During the interviewing and content analysis phase of the research, it became evident that although there is no real gold standard in program evaluation practice, there are certain emerging criteria that are being adopted broadly. More specifically, we observed how most INGOs follow the ‘DAC Criteria for Evaluating Development Assistance’, consisting of: relevance, effectiveness, efficiency, impact and sustainability (OECD, 2014). Indeed, based on our content analysis alone, 62.5% of ToRs name at least four if not five of the DAC criteria (see figure 1), thereby ratifying its prominence in the sector today.

The emergence of these criteria is significant for a number of reasons. For one, it signals that program evaluation has taken a leap beyond the boundaries of the traditional paradigm to embrace other dimensions worthy of measuring in an evaluation. However the particular stress on ‘effectiveness’ is significant, as it appeared more than the other criteria in the ToRs reviewed (see figure 2). Not surprisingly, these findings from the content analysis echo assertions in the literature that “of the five DAC criteria, “effectiveness” best expresses the accomplishment of […] outcomes” (Walker, 2015, p.81).

5.3.5 Program Evaluation: In between paradigms

As can be evidenced by the earlier sections, it is not entirely obvious where program evaluation stands in terms of the overarching paradigms of traditional, outcomes and participatory evaluation.

As witnessed earlier, the donor-derived frameworks for evaluation are still dominant. Although there have been attempts to upgrade the LFA so that its application is more suited to meet the evaluation needs of today (Crawford & Bryce, 2003, p.364), it is still widely regarded as rigid and ineffective (Biggs and Smith, 2003; Smith, 2000 cited in Golini et al, 2015, p.652). The use of these frameworks pin program evaluation practice in the traditional evaluation paradigm where logic and rationality are paramount and tend to be summative in nature – i.e acknowledging end-of-program performance (Rebien, 1997, p. 454). If donors request little else than fulfilling these frameworks to ensure compliance and assert control, then it runs the risk that INGOs treat program evaluation practices as ‘ceremonial and symbolic’ (Arvidson & Lyon, 2013; Cavalluzzo & Ittner, 2004; Gandia, 2011; Gordillo & Andersson, 2004 cited in Mitchell, 2014, p. 608).
Yet as the accountability movement set in, so did changes in these frameworks to reflect a renewed commitment to measuring outcomes and by extension capturing impact (Buckmaster, 1999, p.187). Still, the instances of impact evaluations are far few in between and are often condemned for their poor quality (Mitchell, 2014, p.608). Despite this, these criticisms may be premature as the transition to the outcomes evaluation paradigm is an on-going process. Indeed, it has been cited that donors are adapting frameworks so they move away from fixating on immediate outputs and instead look to understand outcomes (Holvoet & Renard, 2007, p.72).

The emergence of the Theory of Change (ToC) approach in recent years attests to this shift (Carmen, 2010, p.259). The ToC is conceived as “a systematic exploration of the links between activities, outcomes and context” (Actknowledge, 2011, p. 2 cited in Walker, 2015, p.78) and is most commonly employed to assess effectiveness rather than efficiency in a program evaluation (Walker, 2015, p.78). Talk of the ToC surfaced in interviews with INGOs but especially during the content analysis of ToRs. When a key word search was run for words associated to approaches to evaluation, ToC was the second most frequent term after qualitative (see figure 4). This observation is further ratified by the fact that outcomes feature in 70% of ToRs versus outputs appearing in only 45% of ToRs.

What the rise of the ToC demonstrates is that donors have taken some steps to facilitate the induction of the outcomes evaluation paradigm and INGOs are clearly responding. More importantly the emphasis on effectiveness helps bring PM in ID closer to a people-centred paradigm since “outputs are about the program while outcomes are about the participants” (Plantz, Greenway, & Hendricks, 1997 cited in Benjamin, 2012, p.1225).

On this basis, discourse around the current accountability movement would strongly identify with the outcomes evaluation paradigm. However, to say that the participatory evaluation paradigm has held no water would be untrue. The discourse around program evaluation has been far livelier than its ever been, championing the value of understanding process, context and involving beneficiaries (Suárez-Herrera, 2009, p.323). However beyond the mere buzz around the possibilities of participatory evaluation, the uptake has been sluggish. Among the ToRs reviewed, ‘participatory evaluation’ appeared with a frequency of 13%, fourth most significant behind quantitative (see figure 4). Of the INGOs, two placed particular emphasis on their commitment to participatory evaluations (Respondent 2, 2015; Respondent 6, 2015).

Undoubtedly, the state of program evaluation cannot be captured neatly under one paradigm. The tensions, underlying dynamics and good intentions have made it so program evaluation discourse is at odds with practice. On the one hand, the momentum building around evaluation has been positive, with growing recognition of its importance. This is good news as one of the reasons why program evaluation
has largely been inhibited in the past is due to the fact that its importance was undercut and certainly underfunded (Bamberger, Rugh, Church, & Fort, 2004; Council on Foundations, 2002 cited in Carmen, 2010, p.258). A renewed commitment to program evaluation provides some real opportunities for transformation in the sector. With that in mind, the final sub-section seeks to make recommendations.
Chapter 6. Recommendations and Conclusions

6.1 Recommendations

What the data analysis process has revealed is that PM4NGO’s contribution to program evaluation is out-dated and overly simplistic. Meanwhile, program evaluation practices lived out in the sector by INGOs are expanding under the outcomes evaluation paradigm but not without some caveats related to underlying tensions with donors.

In this section a series of recommendations are devised with PM4NGOs in mind. The intention is to upgrade PM4NGOs section titled ‘Monitoring Evaluation and Control Phase’ so that it can more accurately reflect the state and needs of program evaluation practice today, which largely, but not exclusively, sit under an outcomes evaluation paradigm.

• Recognize a multitude of evaluation frameworks

The LFA has no doubt carved out an important space for itself within the imaginary of program evaluation. Despite its drawbacks it clearly has established a degree of permanence in program evaluation. Irrespective, this should not be the only framework presented in PMD Pro. Indeed other frameworks such as RBM and more recently ToC hold more promise. That is to say, they consider outcome dimensions that allow for evaluations to go beyond measuring efficiency and instead prioritize effectiveness.

• Inclusion of DAC Criteria and its guiding questions

That being said, effectiveness is not the only criteria worth considering more earnestly in evaluations. The broad acceptance of the five DAC criteria opens up the possible dimensions of analysis for evaluations. Its inclusion in PM4NGOs would allow program managers and evaluators alike to see the full breadth of possibilities in program evaluation and once more recognizes the importance of going beyond efficiency.

• Unify the approaches and methods behind program evaluation

To really grasp the full scope of program evaluation and all its possible pathways, it would also be advisable to create an organizing chart that details the different types of evaluation, their uses and what methods they work best with. Although there is no updated catalogue of this nature, Rogers & Fraser have given a good framework to develop from as it suggests different kinds of evaluations based on context and nature of the program (2014, p.13-14).
Rather than being strictly prescriptive, this organizing chart could serve as a point of reference to help orientate practitioners on best practices. More importantly, it would create strong added value for the PMD Pro since this kind of information is not readily available elsewhere. In effect, this could help build more visibility and credibility for PM4NGOs as a credible repository of good program evaluation practice. It would also help recognize some of the different approaches that could fall along the spectrum from traditional to outcomes to participatory evaluation (see table 2).

• Tailoring Program Evaluation

Another purpose for creating an organizing chart is that it would allow practitioners to understand that each program can have a unique evaluation approach. This is an important assertion to make because it recognizes, unlike the ethos of PM, that there is no gold standard when it comes to program evaluation. Rather, it is encouraged that practitioners take into account the particular program, context, and beneficiaries to be able to inform their selection of approach. In this process of tailoring, it is fundamental to stress that donor preferences should not dictate the approach entirely and that organizations should feel bold enough to endeavour with more genuine practices like those founded under the participatory evaluation paradigm if it so fits the nature of the program.

• Endorsing learning and not just reporting

Although the guide does recognize the value of iterative learning, this is envisioned more as making changes to the program to keep it on track as opposed to learning from the program’s impact to expand organizational learning and future program performances (Moore, 2003, p. 22 cited in Lee & Nowell, 2014, p.305). Once more a stronger emphasis on learning could help ensure that INGOs are developing a stronger evaluation capacity thus moving beyond simple performance measurements of the traditional era and towards the outcomes evaluation paradigm and even participatory evaluation paradigms (Lennie et al, 2015, p.327).

• Place more emphasis on Evaluation and its counterparts

Lastly, although the conversation focused on evaluation, this is not to presuppose that other functions of this phase are not equally as important. Undoubtedly, monitoring and planning, among others, are fundamental to constructing a cohesive and comprehensive program evaluation strategy. Indeed the guide understands this well by underscoring the importance of connecting the M&E plan to program design documents (Culligan et al, 2011, p.46).
However, the guide also undercuts the value of evaluation as inferred by its claim that evaluation can overshadow other more important functions in the life of a program (Culligan, et al., 2011, p. 16). The PMD Pro can go much further in capturing the importance of evaluation and its counterparts. Drawing clearer linkages between evaluation and other functions and phases of the program will allow this. This is particularly important because it will allow INGOs to more easily embed program evaluation into the wider vision of program management instead of having evaluation relegated to an isolated unit.

6.2 Truth Criteria

In research, the truth criteria are designed to establish the degree to which findings of the study have a degree of credibility. If done well, the evidence collected and subsequent analysis the study should hold up to close scrutiny by peers (Saunders, et al., 2009, p. 156). More specifically, the research should be evaluated on the criteria of validity, reliability and generalizability.

These measures have a great importance in positivist studies since they rely on strict scientific methodology where measurements need to be specific and precise (Saunders, et al., 2009, p. 156; Collis & Hussey, 2014, p. 52). In non-positivist studies and qualitative researches like this one, the criteria for these terms change, since specific and precise measurements are not likely. Instead, the emphasis is placed on the quality and depth of the data collected (Bryman & Bell, 2003, p. 286; Collis & Hussey, 2014, p. 52). For the present study, the interviews conducted with field practitioners and experts in the subject matter helped and the analysis of the ToRs provide this depth and quality.

6.2.1 Validity

Validity refers to the extent to which a specific test is measuring what it was designed to measure (Collis & Hussey, 2014, p.53), meaning there is coherence between what is said to be observed and what it is actually being observed (Bryman & Bell, 2003, p.287). In consequence, the value of data collection is not that it measures observable phenomena but rather that it serves to confirm, and/or negate, propositions asserted by theory (Honan, 2015, p. 60). In this particular study, the validity of data is demonstrable since the interviews and the content analysis of ToRs critically complement and often concur with the literature on how program evaluation practice manifests in the ID field today (Collis & Hussey, 2014, p. 53). It also contrasts the minute impact that PM tools currently have on ID programs, as these have not aided in an increase of success rates of programs within the ID sector. This concurs with the understanding that PM tools should not be equated to a ‘one-size-fits-all’ solution.
6.2.2 Reliability

Reliability is the “accuracy and precision of measurements and the absence of differences if the research is repeated” (Collis & Hussey, 2014, p. 52). Otherwise put, if the same research process were to be carried out by a third party, it should yield the similar results (Bryman & Bell, 2003, p. 288).

When it comes to reliability in an interpretivist study, qualitative data collection is not restricted by the positivist criteria of accuracy and precision of the measurements (Collis & Hussey, 2014, p. 53; Guba & Lincoln, 1994, as cited in Bryman & Bell, 2003, p. 288). When using an interpretivist paradigm the design of methodology should be rigorous enough so that data collection process will always yield authentic results, even if it is conducted during different occasions or by different observers (Collis & Hussey, 2014, p. 53).

Reliability in this study was verified since the same semi-structured interview guide and letter of purpose were issued out in advance to participants; effectively permitting participants interviewed to understand the purpose and scope of the study. For the content analysis, the reliability was secured by ensuring that the same keyword search was systematically and rigorously applied to all the ToRs. In this way, the data produced could be deemed sufficiently reliable for comparisons and analysis to take place. If the study were to be replicated, with similar organizations and similar content to be analyzed, the data collected would yield similar results; that is: a lack of PM contribution to program evaluation and a host of ID-derived program evaluation practices that could inform PM4NGOs.

6.2.3 Generalizability

A work is considered generalizable when the academic and scientific endeavour in one particular environment can be legitimately reproduced and transplanted to another similar environment (Collis & Hussey, 2014, p. 54). The level of generalizability can be considered low in this study given that the research charters a new line of investigation on the topic of program evaluation and in the backdrop of two fields, PM and ID, that are rarely seen together. This is also true due to the small empirical base of evidence found in this study. The limits of which, time and budget, were addressed earlier on. Lastly, generalizability is often an elusive criterion in the social sciences due to the difficulty of predicting human behaviour (MacIntyre, 1990, cited in Lukka & Kasanen, 1995, p. 73).

This being said, this in no way diminishes the assertions and key arguments of the study. It is known that qualitative research is oriented more towards contextual uniqueness (Bryman & Bell, 2003, p. 289), meaning that the results of the study respond to specific moments in time, and specific observations carried out by the
researchers. This is why for the findings to hold in any other context, their needs to be an empirical base (Lincoln and Guba, 1985, p. 316). If the results of the investigation are to be replicated they will be permeated by the next researchers assumptions, limitations and experiences through the observation of the phenomena.

As such, the study may be unable to provide the kind of empirical base for broad claims on program evaluation practices of the sector, however it does carve out niche relevance. Since little research exists to problematize the added value of PM4NGOs, this study offers a point of departure and testable propositions for future research to challenge. Whether or not they become transferable will depend on how future studies confront and qualify their merit. This can either result in propositions being expanded or legitimized further with empirical data or that future research refutes the propositions with counter claims (Collis & Hussey, 2014, p.54).

6.3 Theoretical Contributions and Future Research

While this study has only skimmed the surface of understanding program evaluation in the ID sector and how PM4NGOs can more meaningfully contribute to it, it can assert that the practices currently being proffered by PM lack consideration of the unique characteristics of programs in the ID sector. It has been established that ID programs have marked differences in the way objectives are traced, as well as how and what outcomes are truly measured. In principle, the use of PM tools for evaluation of ID programs seem fitting, due to IDs project based nature. This has been argued by previous research in the field where studies try to show that the adoption of PM tools garner a higher success rate in projects (Golini et al., 2014, p.1183). However, the shift in paradigms within ID from a traditional blue print approach towards a more people centred approach call for a reorientation of these criteria for success. This means aligning with thinking and tools that are more akin to the nature of ID programs and its nuances. What this study has been able to do is capture some of the current modes of thinking and practicing program evaluation in order to meaningfully input into an otherwise stale discourse on how PM can contribute to ID, let alone program evaluation. To truly enrich this line of investigation further, it is worthwhile noting some future research areas.

Although the conversation has centred itself more squarely on how INGOs have experienced program evaluation, this is not representative of the experience faced by local NGOs in developing countries where ID programs are targeted. That is to say, it is well known that the evaluation capacity of local NGOs in developing countries is severely underdeveloped with few to none having an evaluation unit or even staff that can exclusively look at program evaluation (Carman, 2007, p.72).

Another point for potential expansion is including the role of monitoring and planning more earnestly in the scope of future studies. These functions are very much complementary to program evaluation and interact and depend among one another.
Future research that could take on this broader scope could thus provide a more integrated perspective on how to improve PM practices within ID programs. Moreover, the academic discourses also lack a more comprehensive discussion on program evaluation and how this may or not may be distinct from project-level evaluation.

Undoubtedly the intersection between PM and ID remains blurred and poorly understood, and especially so on the topic of program evaluation. Regardless, this study presents and aperture for future research to explore more thoroughly. If anything, this study should have at least underscored the value and potential for cross-sector learning.

6.4 Conclusion

So long as the ID field continues to grow and calls for accountability persist, program evaluation will remain fundamental to reporting and generating learning on how ID programs meet their ‘soft’ development goals. Although program evaluation is fundamentally a management function, PM has not been the one to contribute to program evaluation in any meaningful way but rather the ID sector itself.

At best, PM was able to proffer PM4NGOs and its accompanying guide PMD Pro. Yet their ‘Monitoring, Evaluation and Control Phase’ chapter falls short in capturing the comprehensiveness of program evaluation today. Instead, the guide makes reference to program evaluation frameworks, like the LFA, that are seen as belonging to a traditional era of evaluation, and by extension of PM in ID.

What practice is signalling instead, is a shift towards more people-centred and outcomes oriented paradigms. Under these paradigms, evaluation approaches including the ToC are more equipped to gage performance and lessons learned of ID programs since the emphasis is more squarely on effectiveness and outcomes. It is in this direction that PM4NGOs should shift its attention, driving a greater emphasis on tailored evaluation approaches that are most likely to capture outcomes. Meanwhile, the participatory evaluation paradigm, while not entrenched into practice enough, should serve as a hint to PM4NGOs of the future direction of program evaluation.
References


**Interviews:**


# Appendix 1: Characteristics of Paradigm within Development Evaluation

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Purpose</strong></td>
<td>Focusses on inputs, outputs and deductively tracing the relationship between them</td>
<td>Improves accountability and more effective programming</td>
<td>Encourages consensus and participation among stakeholders by strengthening dialogue and cooperation between stakeholders.</td>
</tr>
<tr>
<td><strong>Use</strong></td>
<td>To record use of funds and demonstrate accountability. There is also an instrumental use where the evaluation follows an evolving logic and rational decision-making process. Feedback to the policy-making level is also paramount.</td>
<td>To measure progress, adjust or redesign programs, develop new initiatives, assess environmental conditions, shape an organisation’s strategic direction and cultivate learning capability.</td>
<td>Two uses: utilization and empowerment. The first is more rational and instrumental, focusing on how technical lessons can feedback into programs whereas the latter is about increasing the number of stakeholders taking ownership in the evaluation and the lessons created.</td>
</tr>
<tr>
<td><strong>Nature of Evaluation</strong></td>
<td>Summative; where providing conclusive and finite information is most important</td>
<td>Formative; where identifying lessons learned for the program is most important</td>
<td>Process oriented</td>
</tr>
<tr>
<td><strong>Contribution to PM in ID Programs</strong></td>
<td>Evaluation is seen as logical and rational and so abides by a project life cycle type model where inputs borne of evaluation contribute to rational decision-making for future interventions and consequently ‘changed practices’. Even though change is of a social nature, it is believe to follow a linear model. Focus at the project level.</td>
<td>Learning tool to give feedback to program managers and facilitate new strategy development. Focus shifted to the program level.</td>
<td>Leveraging communication and strong partnerships, participatory evaluation stimulates a ‘process of collective production of knowledge’ that is co-owned by several stakeholders and more likely to deliver on sustainability. <em>Maintains program level focus</em></td>
</tr>
<tr>
<td><strong>Drawbacks of movement</strong></td>
<td>Evaluation tools are introduced top down from donors; each of them having their own preferred versions, resulting in competing visions about what is worth evaluating from different stakeholders. There is a lack of outcome and process consideration.</td>
<td>The main limitations are a lack of resources and scarce knowledge about approaches’ benefits</td>
<td>Struggles with the context-sensitivity aspect. Considering more stakeholders means more interests have to be juggled. The development community has not been responsive to empowerment angled participatory evaluations; rather they endorse the utilitarian angle; even though this asymmetry comes into friction the main principles of participatory evaluation.</td>
</tr>
<tr>
<td><strong>Examples of tools used</strong></td>
<td>Balanced scorecard</td>
<td>RBM, updated Logical Framework</td>
<td>Fourth generation evaluation, utilization evaluation</td>
</tr>
</tbody>
</table>

*Table 3: Key Characteristics of Paradigm Shifts within Development Evaluation (Rebien, 1997; Buckmaster, 1999; Suarez-Herrera, et al., 2009)*
Appendix 2: INGO Interview Guide

Semi-Structured Interview Guide

Name:
Organization:
Date:

1. What evaluation approaches and methods are most commonly used in the organization?

a. If possible, could you give examples of tools used by the organization to evaluate?

2. Are the evaluation approaches and methods you use introduced by you, the organization, or by donors?

a. If you use methods introduced by donors, which ones?

3. What is the organization's purpose in carrying out an evaluation and what use(s) does it serve?

4. Do you feel differences exist between the way donors see evaluation and the way the organization sees evaluation, if any?

5. Where does program evaluation begin and end?

6. In conducting a program evaluation, what are the criteria normally guiding the evaluation (key words)?

7. What are the strengths of program evaluation in the sector today and areas in which needs to improve?
Appendix 3: Questionnaire on PM4NGOs

This questionnaire is being conducted as part of a master’s thesis research on investigating program evaluation practices of International NGOs (INGOs). The researchers are non-profit practitioners undertaking masters in Strategic Project Management and approach PM4NGOs with the interest of understanding what has been PM’s contribution to program evaluation practices of INGOs. The purpose of this questionnaire is thus to learn about the perspective of PM4NGO and gain insight on how they view their contribution to the international development sector as well as the particular sub-field of program evaluation.

Name of respondent: Edson Marinho
Date: 11/27/2015
Position held in relation to PM4NGOs: Operations Manager

1. When was PM4NGOs formed and why did it come about?
PM4NGO’s was created in 2011. It was created by LINGOs and its partners, to hold the PMD Pro methodology copyright and to develop it.

2. Has its purpose evolved-shifted-expanded since its inception? How so?
Yes. In the last year PM4NGOS is creating a platform to connect all project managers in the development around the world.

3. Has PM4NGOs evaluated its own effectiveness? If so what has it learned?
So far, we are still developing tools to measure effectiveness. It is possible to evaluate the methodology effectiveness, but not the organizational one.

4. Has PM4NGOs measured how many organizations have either adopted the PMD Pro guide or been certified?
Yes, we have case studies and records of the organizations that have adopted PMD Pro. But the records are not confirmed by the organizations and, for that reason, we cannot publicize it. We are contacting the organizations in order to request authorization to publish the information.

We can share that there are organizations from the community-based level to the International Level, such as ChildFund and Heifer International.

5. Regionally speaking, where do most of these organizations come from? (UK, US, Canada, Western Europe, Asia etc)
Most of organizations come from developing countries (Africa, Latin America and Asia). It is possible to see details in the following interactive map:
http://www.pm4ngos.com/we-are-10000-pmd-pros-and-we-want-to-connect/
6. What are the size of these organizations and do they have an international scope to their work? (i.e: INGOs)

Our records are not officially confirmed by the organizations and, for that reason, we cannot publicize it. We are contacting the organizations in order to request authorization to publish the information.

7. With regards to program evaluation, where does the knowledge provided in PMD Pro on this topic stem from? (e.g Does it come from PM? From observed practice in NGOs? another source?)

PM4NGOs is developing the Program Management in Development Guide, which will be launched next year (2016).

8. Has there been an evolution in how PM4NGOs has gone about addressing program evolution since its inception?

Yes. The Program Level Guide is in the final phase of revision - to be published next year.

9. How would the new Program Management in Development Guide be different to PMD Pro?

PMD Pro is focused in the project level, while the Program Guide will focus in the Program Level
Even though the Program Life Cycle, principles and tools will connect with the project level (PMD Pro), they are different and have different purpose (Program level)

10. Is the PMD Pro associated to PM4NGOs?

Yes, it is. PM4NGOs not only has the copyright of the PMD Pro methodology, but most importantly we are responsible for disseminating the methodology around the world.

11. Lastly, if you could comment more specifically about what has been PM4NGOs approach to program evaluation? What are the evaluation practices that are promoted, how in-depth do you go?

As I said, the program guide is still under development. I cannot share the draft guide. Please follow our website or ask me again within 3 to 4 months.
Appendix 4: Interview Guide Prof. Ruggero Golini

This interview is being conducted as part of a master’s thesis research on investigating program evaluation practices of International NGOs (INGOs). The researchers are non-profit practitioners undertaking masters in Strategic Project Management and approach the interviewee with the interest of understanding what has been PM’s contribution to program evaluation practices of INGOs. The purpose of this questionnaire is thus to learn about the perspective of an expert and gain insight on how they view their contribution to the international development sector as well as the particular sub-field of program evaluation.

Name: Ruggero Golini

Organization: University of Bergamo

Date: November 25, 2015

1. What does PM intend to offer the International Development Sector?

2. To what extent do you think PM’s tools have been adopted by INGOs?

3. How would you describe PM’s contribution to ID today?

4. Do you think PM is a field, a sector a discipline or merely a set of tools?

5. Program evaluation is the area of particular interest in our research; to what extent has PM been able to contribute knowledge in this area?

6. How is evaluation perceived in traditional PM?
## Appendix 5: Content Analysis of ToRs (Purposes of Evaluation)

<table>
<thead>
<tr>
<th>ToRs</th>
<th>Learning</th>
<th>Effective</th>
<th>Relevance</th>
<th>Sustainability</th>
<th>Efficiency</th>
<th>Impact</th>
<th>DAC Criteria used? (1/5)</th>
<th>Methodology given</th>
<th>Avg pg length</th>
</tr>
</thead>
<tbody>
<tr>
<td>EveryChild</td>
<td>10</td>
<td>1</td>
<td>3</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>Not as such</td>
<td>N</td>
<td>5</td>
</tr>
<tr>
<td>UNICEF.M</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>4 out of 5</td>
<td>N</td>
<td>5</td>
</tr>
<tr>
<td>Ciff China</td>
<td>5</td>
<td>6</td>
<td>1</td>
<td>3</td>
<td>0</td>
<td>5</td>
<td>4 out of 5</td>
<td>N</td>
<td>10</td>
</tr>
<tr>
<td>Ciff Mexico</td>
<td>0</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>5</td>
<td>4 out of 5</td>
<td>N</td>
<td>13</td>
</tr>
<tr>
<td>Oxfam Bol</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>Not as such</td>
<td>Y</td>
<td>2</td>
</tr>
<tr>
<td>Oxfam Honduras</td>
<td>0</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>Not as such</td>
<td>Y</td>
<td>3</td>
</tr>
<tr>
<td>DRI</td>
<td>0</td>
<td>2</td>
<td>3</td>
<td>0</td>
<td>3</td>
<td>3</td>
<td>3 out of 5</td>
<td>N</td>
<td>3</td>
</tr>
<tr>
<td>Action Aid</td>
<td>10</td>
<td>5</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>3</td>
<td>2 out of 5</td>
<td>N</td>
<td>12</td>
</tr>
<tr>
<td>AIMS</td>
<td>4</td>
<td>4</td>
<td>2</td>
<td>4</td>
<td>2</td>
<td>2</td>
<td>4 out of 5</td>
<td>N</td>
<td>8</td>
</tr>
<tr>
<td>Case for Space</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>4</td>
<td>2</td>
<td>4</td>
<td>5 out of 5</td>
<td>N</td>
<td>4</td>
</tr>
<tr>
<td>Oxfam</td>
<td>1</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1 out of 5</td>
<td>Y</td>
<td>2</td>
</tr>
<tr>
<td>Oxfam Liberia</td>
<td>0</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1 out of 5</td>
<td>Y</td>
<td>2</td>
</tr>
<tr>
<td>Oxfam West Africa</td>
<td>21</td>
<td>7</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>4</td>
<td>1 out of 5</td>
<td>Y</td>
<td>7</td>
</tr>
<tr>
<td>SADC</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>4</td>
<td>5 out of 5</td>
<td>N</td>
<td>4</td>
</tr>
<tr>
<td>Transparency I.</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1 out of 5</td>
<td>N</td>
<td>3</td>
</tr>
<tr>
<td>UN Women</td>
<td>1</td>
<td>9</td>
<td>4</td>
<td>6</td>
<td>6</td>
<td>4</td>
<td>5 out of 5</td>
<td>N</td>
<td>27</td>
</tr>
<tr>
<td>UNHCR</td>
<td>7</td>
<td>5</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>4 out of 5</td>
<td>N</td>
<td>9</td>
</tr>
<tr>
<td>YBI</td>
<td>10</td>
<td>5</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>5 out of 5</td>
<td>N</td>
<td>8</td>
</tr>
<tr>
<td>UNICEF M.2</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>Not as such</td>
<td>N</td>
<td>6</td>
</tr>
<tr>
<td>UNIFEM</td>
<td>3</td>
<td>8</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>5 out of 5</td>
<td>N</td>
<td>10</td>
</tr>
</tbody>
</table>

### Ranked Frequency

<table>
<thead>
<tr>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>3</td>
</tr>
<tr>
<td>4</td>
</tr>
<tr>
<td>5</td>
</tr>
</tbody>
</table>

### OECD DAC Criteria Used as Indicators

- **Effectiveness**: 13
- **Efficiency**: 9
- **Relevance**: 12
- **Impact**: 10
- **Sustainability**: 10
### Appendix 6: Content Analysis of ToRs (Approaches to Evaluation)

#### Approaches to Evaluation

<table>
<thead>
<tr>
<th>Approaches to Evaluation</th>
<th>Theory of Change</th>
<th>Logic Model</th>
<th>Quantitative</th>
<th>Qualitative</th>
<th>Results based</th>
<th>Participatory</th>
<th>Mixed-Methods</th>
<th>Formative</th>
</tr>
</thead>
<tbody>
<tr>
<td>EveryChild</td>
<td>5</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>UNICEF.M</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>CIFF China</td>
<td>4</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Mexico</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Oxfam Bolivia</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Oxfam Honduras</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>DRI</td>
<td>5</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Action Aid</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>AIMS</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Case for Space</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Oxfam Georgia</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Oxfam Liberia</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Oxfam West Africa</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>SADC</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Transparency I.</td>
<td>7</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>UN Women</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>UNHCR</td>
<td>0</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>YBI</td>
<td>7</td>
<td>0</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>UNICEF M.2</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>UNIFEM</td>
<td>10</td>
<td>5</td>
<td>11</td>
<td>15</td>
<td>5</td>
<td>8</td>
<td>7</td>
<td>4</td>
</tr>
</tbody>
</table>

#### Ranking

<table>
<thead>
<tr>
<th>Approach</th>
<th>Ranking 1</th>
<th>Ranking 2</th>
<th>Ranking 3</th>
<th>Tally</th>
</tr>
</thead>
<tbody>
<tr>
<td>Qualitative</td>
<td>8</td>
<td>5</td>
<td>2</td>
<td>15</td>
</tr>
<tr>
<td>Theory of Change</td>
<td>9</td>
<td>1</td>
<td>1</td>
<td>11</td>
</tr>
<tr>
<td>Quantitative</td>
<td>2</td>
<td>6</td>
<td>1</td>
<td>9</td>
</tr>
<tr>
<td>Participatory</td>
<td>3</td>
<td>3</td>
<td>1</td>
<td>7</td>
</tr>
<tr>
<td>Results based</td>
<td>3</td>
<td>2</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>Logic Model</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>Mixed Methods</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td>6</td>
</tr>
<tr>
<td>Formative</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>3</td>
</tr>
</tbody>
</table>

#### Weighted Ranking

<table>
<thead>
<tr>
<th>Approach</th>
<th>Weighted Ranking</th>
</tr>
</thead>
<tbody>
<tr>
<td>Qualitative</td>
<td>2.4</td>
</tr>
<tr>
<td>Theory of Change</td>
<td>2.7</td>
</tr>
<tr>
<td>Quantitative</td>
<td>0.6</td>
</tr>
<tr>
<td>Participatory</td>
<td>0.9</td>
</tr>
<tr>
<td>Results based</td>
<td>0.9</td>
</tr>
<tr>
<td>Logic Model</td>
<td>0.6</td>
</tr>
<tr>
<td>Mixed Methods</td>
<td>0.3</td>
</tr>
<tr>
<td>Formative</td>
<td>0.3</td>
</tr>
</tbody>
</table>

### Tally

<table>
<thead>
<tr>
<th>Approach</th>
<th>3 points</th>
<th>2 points</th>
<th>1 point</th>
</tr>
</thead>
<tbody>
<tr>
<td>Qualitative</td>
<td>1</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Theory of Change</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Quantitative</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Participatory</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Results based</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Logic Model</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Mixed Methods</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Formative</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

79
### Appendix 7: Content Analysis of ToRs (Other Statistics)

<table>
<thead>
<tr>
<th>Indicator</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Outcomes</td>
<td>70%</td>
</tr>
<tr>
<td>Output</td>
<td>45%</td>
</tr>
<tr>
<td>Evidence</td>
<td>70%</td>
</tr>
<tr>
<td>Recommendations</td>
<td>85%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Indicator</th>
<th>Outcomes</th>
<th>Output</th>
<th>Recommendations</th>
<th>Evidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>EveryChild</td>
<td>0</td>
<td>0</td>
<td>6</td>
<td>0</td>
</tr>
<tr>
<td>UNICEF.M</td>
<td>3</td>
<td>2</td>
<td>7</td>
<td>17</td>
</tr>
<tr>
<td>CIFF China</td>
<td>0</td>
<td>3</td>
<td>2</td>
<td>6</td>
</tr>
<tr>
<td>CIFF Mexico</td>
<td>8</td>
<td>0</td>
<td>0</td>
<td>6</td>
</tr>
<tr>
<td>Oxfam Bol</td>
<td>8</td>
<td>0</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>Oxfam Honduras</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>DRI</td>
<td>15</td>
<td>8</td>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td>Action Aid</td>
<td>5</td>
<td>6</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>AIMS</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Case for Space</td>
<td>8</td>
<td>0</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>Oxfam Georgia</td>
<td>8</td>
<td>0</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>Oxfam Liberia</td>
<td>4</td>
<td>2</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Oxfam West</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Africa</td>
<td>3</td>
<td>4</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>SADC</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Transparency I.</td>
<td>6</td>
<td>7</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>UN Women</td>
<td>0</td>
<td>0</td>
<td>5</td>
<td>3</td>
</tr>
<tr>
<td>UNHCR</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Colombia</td>
<td>9</td>
<td>1</td>
<td>6</td>
<td>0</td>
</tr>
<tr>
<td>YBI</td>
<td>1</td>
<td>0</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>UNICEF.M 2</td>
<td>7</td>
<td>6</td>
<td>6</td>
<td>3</td>
</tr>
<tr>
<td>UNIFEM</td>
<td>14</td>
<td>9</td>
<td>17</td>
<td>14</td>
</tr>
</tbody>
</table>