

INGO Performance Management Practices
And the Need for Accountability

Bradley Conlin

Thesis proposal submitted
In partial fulfillment of the requirements
For the Telfer Master of Science (MSc) in Management Program

Telfer School of Management
University of Ottawa

© Bradley Conlin, Ottawa, Canada, 2015

Abstract

This study provides a comprehensive look at the current state of best practices in performance appraisal systems in the public and private sector. From this, the main objective for this research paper is to investigate the reasons why INGOs are not currently able to properly implement a 'best practices' based performance appraisal system. Furthermore, this paper will provide both future research questions, as well as practitioner based recommendations. In addition, it will examine the need for higher levels of accountability in INGOs through appropriate performance appraisal systems.

Table of Contents

1. Introduction	1
2. Accountability in INGO Management	3
3. Literature Review on Performance Appraisals	8
3.1 Organizational Justice and Equity Theory	9
3.2 Judgmental Purpose	10
3.3 Developmental Purpose	12
3.4 Assessment Criteria	13
3.4.1 Inputs vs. Outputs Criteria	14
3.5 Instrumentation	17
3.6 Biases and Rater Training	18
3.7 Coaching	20
3.8 Sources of Appraisal	21
3.9 Frequency	23
4. Methodology	26
4.1 Design	26
4.2 Sample	26
4.3 Procedure for Data Collection	27
4.4 Procedure for Analysis of Results	29
5. Results	31
5.1 Overview	31
5.2 Research Question #1 – Accountability	31
5.3 Research Question #2 – Private Sector Best Practices	35
5.3.1 Assessment Criteria.....	36
5.3.2 Instrumentation.....	39
5.3.3 Biases and Rater Training	42
5.3.4 Coaching.....	44
5.3.5 Sources of Appraisal.....	47
5.3.6 Frequency	48
6. Discussion	51
6.1 Summary and Theoretical Implications	51
6.1.1: Finding #1: A Successful Performance Appraisal System is Greater than the Sum of its Parts	52
6.1.2: Finding #2: Uncritical Accountability to Donors Raises Potential Liability	55
6.1.3: Finding #3: Introduction of a Private Sector Tool to a Nonprofit Organization Potentially Represents a Culture Change	56
6.1.4: Finding #4: The INGOs’ Cultural Value for Human Resource Development is an Asset...	57
6.2 Practical Implications	57
6.3 Limitations and Directions for Future Research	59
6.3.1 Limitations	59
6.3.2 Directions for Future Research	59

7	References	61
8	Appendix A	68
9	Appendix B	71
10	Appendix C	73

1. Introduction

INGOs refer to “international non-governmental organizations” (Roberts, Jones & Frohling, 2005), organizations that serve a wide range of socioeconomic purposes. These organizations have proliferated over the past years (Vays, 2010). With over \$100 million of government aid being given to Canadian INGOs every year, the need for INGOs to be accountable to their donors is becoming more apparent (CIDA, 2006, O’Sullivan, 2010).

In the 2007 book “The Management of Non-Governmental Development Organizations”, David Lewis identifies numerous areas where both research and industry are lacking knowledge of how human resource challenges affect performance in INGOs (Lewis, 2007). More recently, research suggests that international NGOs face barriers to applying private/public sector human resource management “best practices” due to their volatile context (O’Sullivan, 2011). Contextual challenges include having to deal with partners in time zones that are up to twelve hours apart, overseas partners being situated in dangerous and unreliable territories, lack of qualified applicants in selection pools, etc. (O’Sullivan, 2010).

Given the apparent challenges of managing performance in INGOs, and the need for accountability, but limited research in this area, the purpose of this study is twofold. First, it seeks to justify the importance of performance appraisals for INGO employees by reviewing pressures for accountability. Second, it provides an in

depth review of the private sector literature on best practices for the design, implementation and utility of performance appraisal systems (Noe, 2005), with the aim of exploring the extent to which existing performance appraisal theory may be applied to the INGO context, given the accountability pressures that exist.

It is from these two main research purposes that the research questions for this paper have been formulated. The first question will be used to reflect on the perceived need for accountability in the organization, while the second question will be used to reveal information concerning private sector HR best practices. Therefore, the research will be driven by the two following questions:

- 1) How, if at all, is the need for accountability in INGOs reflected in how they conduct appraisals?
- 2) How are the best HR practices for performance appraisals (as advocated in the private sector management literature) actually being applied to performance appraisal systems in INGOs? And if/when they do deviate, why?

Furthermore, the first two sections of this paper address the dual purposes noted above. The remainder outlines the method and results of an empirical study, which investigated these issues. The thesis concludes with a discussion of the theoretical and practical implications.

2. Accountability in INGO Management

INGOs exist for “diverse socioeconomic reasons, ranging from global environmental concerns (e.g., Greenpeace)... global human rights concerns (e.g., Amnesty International), or... humanitarian health assistance in times of crisis (e.g., Medecins San Frontieres)” (O’Sullivan, 2010, p. 421; Cavill & Sohail, 2007). While most people think of an INGO as having one particular function or goal, “in reality, many focus on a variety of socioeconomic aid objectives, such as the Red Cross Movement, which consists of a “Federation” that deals with international development, an “International Committee” that deals with health/logistical issues in humanitarian emergencies/conflict zones, and various national societies (e.g., the Canadian Red Cross) which deal with domestic preparedness and disaster management” (O’Sullivan, 2010, p.421). This is because while many INGOs deal primarily with one focus (i.e. health), they typically need to aid the foreign country in several other matters in order to achieve their goals (i.e. transportation, electricity, infrastructure, etc.) (Cavill & Sohail, 2007).

A significant amount of funding and government support is given to INGOs (O’Sullivan, 2010). In Canada alone, international development NGOs receive over \$100 million annually in financial aid via Canada’s Overseas Development Assistance (ODA) budget (CIDA, 2006). Furthermore, the amount of funding that Canadian development INGOs receive may more than double, if Canada increases its ODA rate of 0.32% of GNI to its committed level of 0.7% of GNI in the future (O’Sullivan, 2010, CIDA, 2006).

O'Sullivan (2010) highlighted the debate between those who question whether there should be more funding given to a seemingly corrupt misuse of public money (Labelle, 2007) and those who argue the need for an even more drastic increase in funding for these organizations as a way of circumventing corrupt local governments (Sachs, 2007). While it is not the place of this paper to decide these issues, it is clear that there is a growing urgency for increased accountability by these INGOs (Cavill & Sohail, 2007).

Accountability can have multiple meanings to many people, depending on the context and the purpose for its use (Cavill & Sohail, 2007). The questions raised about INGO accountability stem, in part, from a lack of clarity about the social impact of their achievements. For example, number of clients served is a common metric, but it fails to indicate how effectively those clients were served (Cavill & Sohail, 2007). That is to say, quantity of clients served is being reported, but not the actual quality of those services. Cavill and Sohail (2007) have identified multiple reasons for the need for greater accountability in INGOs, which have adapted into the following table.

Table 2: Pressures for INGO Accountability (Adapted from Cavill & Sohail, 2007)

<p>Rising political visibility of INGOs and growth of the sector</p>	<ul style="list-style-type: none"> • In recent years, UK-based INGOs have shifted their focus from service delivery to building local capacities and promoting participation for local problem solving, to rights-based work, and most recently to lobbying and advocacy for poor people and monitoring government. • The growing influence of INGOs over decision makers and policy makers on behalf of marginalized constituencies has led to concerns about the power of INGOs.
<p>A Crisis of Legitimacy</p>	<ul style="list-style-type: none"> • A widespread crisis of the legitimacy and governance of global corporations and government has also affected the INGO community. • INGOs are perceived as having lost their radical origins (Wallis, 2004).
<p>Accusations of corruption from the press/international policy makers</p>	<ul style="list-style-type: none"> • INGO misbehavior has been highlighted in terms of misuse of funds, the abuse of power over vulnerable people, dishonesty in fund-raising/advocacy.
<p>Demand for good governance of international development practice</p>	<ul style="list-style-type: none"> • INGOs are expected to be better able to meet donor demands for professional governance standards (especially with regard to communication and reporting). Examples here include the collapse of Children’s Aid Direct, one of the largest humanitarian INGOs in Britain, and allegations of sexual exploitation, and abuse of refugees by humanitarian workers in West Africa. • The potential of accountability as a tool to help an organization to promote better performance has been recognized. Accountability enables an organization to learn from its mistakes and successes, as well as to reduce information asymmetries between INGO and client communities.

In addition to these pressures for INGO accountability Cavil and Sohail (2007) also describe the four directions for INGO accountability, which need to be addressed at global and local levels.

Table 3: Directions of INGO Accountability (Adapted from Cavil & Sohail, 2007)

Direction of Accountability	Stakeholders Involved	Explanation
Upward Accountability	Donors, Funders, Boards of Trustees, and Host Governments	Ensuring that INGOs deliver value for the money and meet target deadlines
Downward Accountability	Individuals or local beneficiaries (including local NGOs) and those to whom they provide services or on whose behalf they speak	Ensuring that quality service is provided to clients and feedback on service quality is obtained from clients
Horizontal Accountability	Peers and fellow professionals	Ensuring that peers meet shared values and standards to uphold the standards and reputation of the sector.
Inward Accountability	Staff	Accountability toward staff who are adhering to their organizational mission and values by working in accordance with normative expectations.

This close examination of INGO accountability has revealed important distinctions from how the notion of accountability is treated in the private sector. In private sector organizations, any waste is that of the company's own profits, whereas in an INGO, the waste is that of government funding and charitable donations. In addition, those who fund international development projects (e.g. national donors, host governments, citizens from developed economies) are not the same as those who directly benefit from it (e.g. local recipients of the services). This is a stark difference from the private sector because, in most private sector transactions, the customer pays the organization for a product or service that they require, which enables the customer to more easily assess the services or products being purchased. By contrast, in an INGO interaction, there are four separate stakeholders with an interest in the transaction: payee (funder), organization, host government, and

customer (recipient/beneficiary who benefits from the goods/services) (Lewis 2007, p. 35-41). This makes it more difficult for payees to assess the goods and services that they purchase, as they will not be receiving these directly.

Clearly, with the increased need for funding, and competition for said funding (Cavill & Sohail, 2007), a need for management practices that support accountability has increased dramatically in recent years (O'Sullivan, 2010; Ramia, 2003). Thus, it is important that the performance appraisal system being used by an INGO reflect all of these four paths of accountability. However, precisely how each of these accountability directions is prioritized in practice remains an under-researched issue.

3. Literature Review on Performance Appraisals

Performance appraisals are possibly one of the most important aspects of a manager's job (DeCarlo & Leigh, 1996; Larson, 1984; Morris, Davis, Allen, Avila & Chapman, 1991) because individual performance improvement plans can help the organization to optimize performance. Indeed, many researchers consider the performance appraisal to be one of the most important tools accessible to managers (Boswell & Boudreau, 2000; Pettijohn et al., 2000; Pettijohn et al., 2001a; 2001b).

The majority of top 1000 performing companies in the world use some form of performance appraisal (Smith et al., 2000). The irony is that while the goal of a performance appraisal can often have a negative effect on employee morale (Smith et al., 2000), its goal is to be not only judgmental, but also developmental: to help develop employees by enabling them to identify their negative and positive actions (Smith et al., 2000). However, if not well managed, the performance appraisal is a performance management tool that can adversely change the climate of the workplace for weeks before, and even after the appraisal event has occurred (Pettijohn et al. 2001; Smith et al., 2000).

In order to fully appreciate how a reliable and valid performance appraisal system can contribute to accountability, this review will examine what the management literature considers to be "best practices" for performance appraisal systems. It will also examine why these are considered to be best practices. Namely, at the core of an accountable performance appraisal system there must be organizational justice.

Therefore, the first section of this review will start by describing the theoretical lens of organizational justice. The remaining sections devote attention to the best practice elements of a performance appraisal system such as: judgmental and developmental purposes, assessment criteria, frequency of appraisals, rater training, coaching during the feedback session and sources of appraisal.

3.1 Organizational Justice and Equity Theory

Organizational justice incorporates notions of both procedural and distributive justice (Karriker & Williams, 2009). Procedural Justice refers to the perceived fairness of the organizational processes that are used to reach decisions (Karriker & Williams, 2009). Distributive justice refers to the “perceptions that individual reward levels are consistent with employees’ contributions to the organization” (Heneman, H. G., 2007, p. 391). Distributive Justice is thus implicitly dependent on the justice of the procedures used to assess employee contributions.

The concept of organizational justice is grounded in Equity Theory (Karriker & Williams, 2009). Equity Theory states that individuals assess the outcomes received from their performance appraisal (e.g. compensation, praise, etc.) in relation to their input (e.g. hours worked, sales made, etc.) (Karriker & Williams, 2009). A sense of inequity arises when an individual’s perception of what outcomes are deserved for their input does not match with their expectations (which can also be rooted in a mistaken “referent other”). When employee perceptions of either procedural or

distributive justice are negative, research has shown it can adversely affect the organization in a number of ways, including impeding organizational citizenship behaviour (Karriker & Williams, 2009).

3.2 Judgmental Purpose

Judgmental feedback is the direct feedback about performance during the evaluation period (Pettijohn et al., 2001). Judgmental feedback is focused on the quality of actions taken in the past, in comparison to the developmental feedback which is future focused (Pettijohn et al., 2001).

Judgmental feedback from a performance appraisal can help the organization on a number of different levels (Pettijohn et al., 2001). First, at the individual level, when performance appraisals are perceived to have been conducted in a procedurally just manner, they help increase employee motivation, performance and organizational citizenship behaviours (OCB)(Pettijohn et al., 2001; Boswell & Boudreau, 2000; Smith et al., 2000). These employee outcomes arise when feedback from the performance appraisal helps to decrease the employee's level of role ambiguity (Smith et al., 2000). Research has shown that decreased role ambiguity not only contributes to increased employee performance, motivation and OCBs but also increases employees' job satisfaction and decreases employee absenteeism and turnover (Li & Bagger, 2008).

Second, although this paper focuses on individual-level performance assessments, the significance of individual performance appraisals can be better appreciated by recognizing how judgmental feedback to the individual also contributes to organizational performance. Knowledge of employee performance also yields direct organizational level benefits. It enables firms to assess the effectiveness of their company's HR policies and maximize their human resource management investment (Pettijohn, Parker, Pettijohn and Kent 2001; Smith, Harrington and Houghton, 2000). Crucially, the judgmental outputs of performance appraisals also enable the organization to distribute justice appropriately: they allow an organization to recognize and reward its top talent via salary increases, merit pay increases and other forms of total compensation (Beem, 2001; Boswell & Boudreau, 2000; Decotiis & Petit, 1978; Pettijohn, Pettijohn & Taylor, 2000; Pettijohn et al., 2001; Pettijohn, Pettijohn, Taylor & Keillor, 2001; Smith et al., 2000).

However, when undue emphasis is placed on the judgmental aspect of the performance appraisal, it can create discomfort for both the employee and the rater (Smith et al., 2000). To mitigate the risk of these negative consequences, appropriate steps must be taken during the judgmental component. Namely, properly communicating to employees the rationale for the appraisal and allowing employees to share their own feedback about the appraisal process can enable organizations to realize significant increases in employee acceptance of the appraisal system (Boswell & Boudreau, 2000; Smith et al., 2000; Larson, 1984). In

addition, attending to the developmental aspect of the appraisal is also important, as the next section will describe.

3.3 Developmental Purpose

Judgmental information about the recent performance of their employees (Boswell & Boudreau, 2000; Pettijohn et al., 2001) inevitably highlights areas where employees need further development. Developmental feedback typically takes the form of action plans for employees. Action plans may consist of formal training objectives and/or goal-setting action plans to develop on-the-job competencies (Pettijohn et al., 2001). When employees' perceive that the performance appraisal is also being used for developmental purposes, it improves their acceptance of the performance appraisal system as a whole, as well as their approval of the person conducting the appraisal (Boswell & Boudreau, 2000).

Use of the appraisal for developmental purposes can also enhance the organization's group dynamics (Boswell & Boudreau, 2000), which can improve organization climate as well. Although this paper is not directly concerned with organizational-level variables such as climate, understanding how developmental feedback contributes to such organizational-level benefits helps to underscore the importance of the developmental aspect of performance feedback too. Namely, it suggests a strategy for improving the reputation of this often-misaligned HR tool (Boswell & Boudreau, 2000; Smith et al., 2000). When creating a new performance

appraisal system (or when an organizational culture already views the current performance appraisal system negatively), it is advisable to include a developmental aspect to the appraisal to improve perceptions of organizational justice (Boswell & Boudreau, 2000; Smith et al., 2000).

3.4 Assessment Criteria

The criteria used to assess performance are fundamental to the effectiveness of the appraisal (Boswell & Boudrea, 2000, Pettijohn et al., 2001a, 2001b). There are two key reasons for this.

First, if the criteria being assessed do not fully reflect the activities performed by the employee, one of two problems will occur: criterion deficiency or criterion contamination (Boswell & Boudrea, 2000). Criterion deficiency means employees will not be evaluated on what they are implicitly expected to achieve. Criterion contamination means employees will be evaluated on results that are not truly within the context (or control) of their job description. In either case, emphasis of where employees need to focus efforts will not correspond to their job and business function (Boswell & Boudreau, 2000; Smith et al., 2000), and individual motivation (and by extension, performance) will not advance in the desired direction. If that occurs, then all the potential employee-level benefits of a performance appraisal system (as reviewed in the previous section) will be severely lessened if not completely negated (Boswell & Boudreau, 2000; Morris et al., 1991; Smith et al., 2000). Poor criteria can adversely influence the degree to which the evaluation

process will be accepted by both employees and their managers (DeCarlo & Leigh, 1996; Boswell & Boudreau, 2000; Morris et al., 1991; Smith et al., 2000; Pettijohn et al., 2000a, 2000b, 2001a, 2001b). Conversely, the more relevant the appraisal criteria are to the employee's position in the organization, the more the employee and manager will feel the appraisal is valid (Smith et al., 2000). This phenomenon is documented in the Organizational Justice literature (Karriker & Williams, 2009), as noted earlier.

The second problem with inappropriate criteria occurs at the organization level, because the information being gathered and sent upward to management will be misleading (Larson, 1984; Morris et al., 1991; Pettijohn et al., 2000; Pettijohn et al., 2001a, 2001b). Inaccurate assessments of the organization's overall performance can jeopardize the competitive strategic advantage that the performance appraisals are meant to add to the organization. Therefore, to avoid adverse organizational-level implications, it is critical to use appropriate individual-level criteria.

3.4.1 Inputs vs. Outputs Criteria

Although the importance of a fully representative set of appraisal criteria is generally well recognized, the preferred nature of those criteria is not nearly as well agreed upon in the academic literature (Pettijohn et al., 2001a). Researchers will usually divide evaluation criteria into two main categories. These categories may have many different labels (e.g. subjective/objective, qualitative/quantitative, input/output, etc.). They are, however, most often referred to as inputs and outputs

(Pettijohn et al., 2001a). Inputs are generally “soft”, more subjective characteristics, such as; personal qualities, ideas, strategies used and activities performed. Conversely, outputs tend to be more objective in nature (e.g. Using “harder” criteria measures such as sales volume, customers served, quantities of work done) (Pettijohn et al., 2001a).

Research has shown that there are two reasons that output (hard/objective) criteria are generally more favoured by organizations and managers (Pettijohn et al., 2001a). The first reason has to do with the easily measurable nature of the output criteria. When managers conduct annual evaluations they find it significantly easier to look up the records that are numerically based than to objectively assess the input effort. The second, and somewhat related, reason that input criteria may be less favoured by organizations is due to the difficulty in assessing the growth of an employee’s inputs over time (Pettijohn et al., 2001a, 2001b; Smith et al., 2000). For example, how does a manager measure a shift in an employee’s strategies to increase sales, if not by basing this assessment on objective output criteria? Although it may be easy for a manager to give praise to an employee who is showing higher levels of motivation, it is very difficult to quantify the value of that effort (Smith et al., 2000). The inherent difficulty in measuring inputs has caused the standard for appraisal criteria to be the use of outputs (Pettijohn et al., 2000a, 2000b; Pettijohn et al., 2001a, 2001b; Smith et al., 2000).

On the other hand, contrary to what *organizations* prefer, *employees* have been found to favour evaluations based predominantly on inputs rather than outputs (Smith et al., 2000). That is, despite their recognition that outputs are easily and objectively measurable, most output criteria are disliked because it risks offending employee notions of procedural justice and they are regarded as somewhat out of the control of the employee (Pettijohn et al., 2001a; Smith et al., 2000).

The dilemma over whether to opt for predominantly input vs. output criteria becomes further complicated when the judgmental information from the appraisal forms the basis for compensation or career decisions. When using intangible input criteria to determine a tangible reward, it can be very difficult to preserve employee perceptions of distributive justice (Smith et al., 2000). Thus, the paradoxical reality is that an input based assessment can help to increase employee perceptions of procedural justice and acceptance of the appraisal, but it can also decrease their perceptions of distributive justice. Output based assessments can have the opposite effect (i.e., increasing distributive justice, while decreasing procedural justice and appraisal acceptance)(Smith et al., 2000). Input criteria have also been found to be positively correlated job satisfaction (how content an individual is with their job) (Smith et al., 2000), while output criteria have been positively associated with job performance (the level at which an employee is able to complete their job tasks) (Smith et al., 2000). Thus organizations must grapple with identifying a contextually appropriate balance between the two.

These two types of assessment criteria have been a matter of debate for years. The output-based approach has been used the most in practice (Pettijohn et al., 2001a; Smith et al., 2000), as its objectivity and ease of implementation makes it significantly more legally defensible. However, research on hybrid approaches (i.e. using both input and output criteria) has supported the merits of a hybrid approach (Pettijohn et al., 2001a). Indeed, the weight of research on the type of criteria that should be used in performance appraisals is more supportive of a hybrid approach than either an exclusively input or exclusively output approach (Boswell & Boudreau, 2000; Morris et al., 1991; Pettijohn et al., 2001a; Smith et al., 2000).

The challenge of balancing perceptions of procedural and distributive justice when using hybrid criteria can be partly mitigated via attention to instrumentation (Boswell & Boudrea, 2000). Instrumentation will be discussed in the next section

3.5 Instrumentation

As with anything that has been around for centuries, the performance appraisal instrument has seen its fair share of alterations and changes. The first documented formal evaluation took place in North America in 1813, by an Army General who had submitted an evaluation of all his men to the War Department (Wiese & Buckley, 1998). These original evaluations used forms that have since been termed as “global ratings and global essays”, which are rating forms that provide a general comment, with little to no distinction between rating dimensions (Wiese & Buckley, 1998; DeVries, Morrison, Shullman & Gerlach, 1981). This type of rating would provide

general feedback on the overall performance of the employees in mention, with typical ratings including “outstanding”, “satisfactory” and “needs improvement” (Wiese & Buckley, 1998; Devries et al., 1981).

These standards eventually evolved to enable more accurate and useful approaches to the evaluation system. Table 1 in appendix A highlights some of the trends seen in the tools used to assess performance over the last two centuries. As Table 1 indicates, the state of the art in performance appraisal instrumentation now advocates use of a combination of MBO (for judgmental purposes of evaluating outcomes) and Behaviorally anchored rating scales (BARS) (for developmental purposes of evaluating inputs). That is to say that a proper performance appraisal instrument should permit customization for each position in the organization to help ensure that the performance appraisal tool used permits measurements of both input and output criteria. This approach will allow for not only reporting of judgmental data, but also help with development of the employee, and secure greater reliability and validity.

3.6 Biases and Rater Training

With any situation where best practices are implemented there will always be some degree of human error. As Table 1 in appendix A has shown, performance appraisal tools have evolved over time to become less subjective, with the intention of reducing the possibility for human error. Indeed, research has shown (with some mixed reviews) that ratee sex, age and gender affect performance outcome

insignificantly (Bretz, Milkovich & Read, 1992). Nevertheless, there still remain human error issues involved in performance appraisals that may occur independent of rater/ratee personal demographic characteristics (Bretz et al., 1992).

One example is the halo/horns effect that occurs when a rater tends to be naturally too easy or hard on an employee (Bretz et al., 1992). The halo effect refers to when employees receive an inaccurately high rating on their performance appraisal (Bretz et al., 1992). The horn effect is the opposite, where the rater will tend to inaccurately rate the employees lower than appropriate (Bretz et al., 1992). In either case, the rater may be unaware of these biases, as it is a natural outcome of their personality and the training they receive (or failed to receive) (Bretz et al., 1992). More problematically, these effects can occasionally be introduced purposefully if the rater feels that the actual performance ratings may reflect badly upon his/her own management abilities (and may risk sending an unflattering message to upper management) (Bretz et al., 1992).

While some degree of human error is seen as unavoidable, proper rater training can help to increase reliability and validity of the tools being used (Bretz et al., 1992). This occurs for two reasons. First, rater training can enable raters to gain a better understanding of the appraisal tool, which will allow for more accurate and reliable outcomes (Bretz et al., 1992)¹. Second, the proper training of raters will strengthen

¹ This enhanced awareness is a double edged sword, however, as unscrupulous raters could then be better informed of how to use the tool for personal gain, and can intentionally change the evaluations to meet their needs better (Bretz et al., 1992).

rater's observational skills. These higher-level observation skills will increase the rater's self-efficacy towards conducting the performance appraisals, which in turn will allow for more accurate appraisals (Bretz et al., 1992). This can also lessen the rater's own skepticism that the tool might need to be misused in order to keep the employee and management satisfied (Bretz et al., 1992). When rater understanding and self-efficacy are increased in this manner, it has been found to reduce employee dissatisfaction with the appraisal meeting. This is because employees' perceptions of procedural justice increase along with a greater understanding of the appraisal itself (Bretz et al., 1992).

In addition to receiving training on how to gather information to assess ratee performance and then provide a judgment on performance, raters can also potentially receive training on how to meet the developmental objective of the appraisal (Bretz et al., 1992). This developmental objective is achieved by the performance appraisal discussion, and is ideally expressed using a coaching approach, which is reviewed in the next section.

3.7 Coaching

A coaching approach is especially important when a pay increase or promotion is not possible and/or warranted (Bretz et al., 1992). The coaching process during the performance feedback discussion should ideally provide employees with a better understanding of their actions and the outcomes (Bretz et al., 1992). Even when an increase in pay or a promotion occurs, research indicates it will leave the employee

happy, but the employee still may be dissatisfied with the appraisal discussion if it was done improperly or conveyed in an ineffective manner (Bretz et al., 1992). When an employee feels that procedural or distributive justice is lacking, (e.g. that they either do not deserve the awards given to them, or even worse, that they feel they deserve more), they usually want a concrete explanation. By contrast, when these two forms of justice appear to be present in the explanation of how everyone is being assessed, employees tend to be more accepting of receiving the awards that they receive.

Research has shown that there are three dimensions to effective coaching feedback: being supportive, emphasizing improvement and discussing rewards (e.g. pay and advancement) (Bretz et al., 1992). By providing feedback along these three dimensions, the employee will leave the meeting far more satisfied than if this approach were not used (Bretz et al., 1992). This is because employee's perceptions of procedural justice increase when they gain a better understanding of how and why they are perceived as they are by the rater (Bretz et al., 1992; Heneman, H. G., 2007).

3.8 Sources of Appraisal

Originally, performance appraisals were completed only by the supervisor of the employee, who would use whichever performance appraisal instrument their organization had available (Wiese & Buckley, 1998). However, this traditional

approach raises issues about the reliability of the measurements, because only one perspective is being used (Wiese & Buckley, 1998).

In more recent years, organizations have used what is called 360-degree feedback. Research has shown that at least 40% of organizations have been using this approach (Antonioni & Park, 2001). With 360-degree feedback, individuals are assessed by three or four sources: they assess themselves; they receive feedback from their peers, and their supervisor(s); and, if they are a manager, they also receive feedback from their subordinates (Antonioni & Park, 2001). However the 360-degree feedback approach is done primarily for developmental purposes, because, when used for judgmental (and reward) purposes, promotions and pay increases are incentives enough to potentially skew the self and peer reported assessments (Antonioni & Park, 2001). Moreover, upward appraisals are only recommended in contexts of many subordinates, to avoid socially desirable response bias that might arise due to a lack of anonymity among a small group of subordinates (Antonioni & Park, 2001).

When used appropriately, this approach has several benefits: First, self-assessment enhances employee acceptance of the appraisal process (Beem, 2001; Boswell & Boudreau, 2000; Decotiis & Petit, 1978; Pettijohn, Pettijohn et al., 2001; Smith et al., 2000). Second, use of additional sources helps to counteract biases, such as the interpersonal effect of leniency, also known as liking (Antonioni & Park, 2001). "Liking is an emotional reaction (positive, neutral, or negative) to a specific person.

Even though rater affect is irrelevant to evaluating an individual's performance of given tasks, the fact that a rater likes or dislikes a ratee can influence the ratings and consequently determine that rater's evaluation of performance" (Antonioni & Park, 2001, p.480). Third, having multiple sources of rating input not only mitigates the possibility of accidental biases or maliciously skewed rankings, but it also helps the employee gain a better understanding of how they are perceived in the organization and in their role (Antonioni & Park, 2001). This further enhances the potential for employee acceptance of the appraisal process.

3.9 Frequency

The last key element of a high performance appraisal system is the frequency with which appraisals are conducted (Pettijohn et al., 2001a). Research has indicated that frequency plays a vital role in both the perceived validity of the appraisal, and subsequent ratee acceptance (Pettijohn et al., 2001a). Deciding upon an appropriate frequency of performance feedback meetings with employees has three key benefits:

First, although annual reviews are still common in many organizations, decreasing the performance assessment period allows for easier assessment by the manager (Pettijohn et al., 2001a), because it enables the manager to assess performance more accurately (Pettijohn et al., 2001a). Often, managers can only remember the past 3-4 months effectively, with the remaining 8-9 months having far less accuracy (Pettijohn et al., 2001a). For this reason, many organizations increasingly use a bi-

annual performance period, as they find it an easier timeframe in which to give more accurate feedback, and thus a more practical time period (Pettijohn et al., 2001a).

The second way that a shorter assessment period helps an organization is with employee acceptance. Having bi-annual assessment meetings will improve the employees' acceptance of the feedback given (Pettijohn et al., 2001a). This is because it also enables the employee to recognize a higher level of judgmental accuracy, and both parties can remember the key points in the previous feedback session more easily.

The third, and most important way in which an increased frequency will help the employee is with an increased amount of opportunities to obtain clarification about their job requirements and their fulfillments of these (Pettijohn et al., 2001a). Thus, by correcting incorrect behaviour and rewarding positive behaviour more than once a year, it will allow for better understanding of the employee's job, as well as reduce the chance that bad habits might set in (Pettijohn et al., 2001a). This will also strengthen the developmental experience of the assessment as the employee and manager will have more of a conversational approach to discussing performance, rather than strictly focusing on judgmental feedback (Pettijohn et al., 2001a).

Although an increase in assessment frequency is likely to be beneficial, there is a need to strike a balance between what is most effective against what is most

practical for any specific organization (Pettijohn et al., 2001a). This is why many organizations will encourage informal assessment and coaching conversations between management and employees (Pettijohn et al., 2001a). This approach will allow for a more consistent feedback loop, without the need to commit the significant time of a formal feedback procedure (Pettijohn et al., 2001a). The only precaution here is to ensure that the feedback does not become too frequent as it can be seen as micro managing and/or indicative of a lack of trust (Pettijohn et al., 2001a). There is no identified optimal time frame for the frequency of informal assessment, as specific jobs and specific employees may require more or less than another (Pettijohn et al., 2001a).

4. Methodology

The research questions were explored through a qualitative approach, which built on the knowledge gained from the preceding literature review (Creswell, 2007). A qualitative research approach was chosen because qualitative designs should be used when there is a need for information gathering in a natural setting and there is a need for a complex understanding of issue(s) at hand (Creswell, 2007, p.37). With very little research done on the topic of this study, it is essential for a high level of understanding and themes to be created around this topic (Yin, 2009).

4.1 Design

This study used a single case study approach, which involves the study of an issue explored through one or more cases in a bounded system (Creswell, 2007, p.73). More specifically this study used an instrumental case study approach, whereby the researcher focuses on a specific topic, and then chooses a bounded case which allows for a rich, deep understanding of the topic (Creswell, 2007, p.74).

4.2 Sample

The organization chosen is an INGO that the author is personally familiar with, and is one in which performance management has been identified as a problem. This organization is a prominent INGO in the education development sector. It has all of its operations located in Ottawa, Canada. However, the organization uses partners from the countries receiving aid to co-implement the education programs created in

Ottawa. This organization is just over 50 years old, and averages around 18-20 employees.

Due to the nature of the organization, each employee was interviewed, allowing for a “purposeful maximal sampling” which is viewed as a superior approach to sampling when doing a single case study, as it allows for multiple perspectives on the same point (Creswell, 2007, p.74). Thus the sample size consisted of 21 employees (3 managers and 18 staff). The first interview was conducted with the organization’s CEO (as this is the only strategy based position in the small organization), and the interview protocol was adapted to that individual’s position. The questioning of management was intended to help gather the views of those with a more strategic perspective. The subsequent interview(s) were conducted with non-managerial employees. These two categories of interviewees were chosen to provide potentially contrasting views of the performance appraisal system within the organization, which would not only help to increase the understanding of these complex topics being studied but also help to increase the validity of the results gained (Yin, 2009, p.108).

4.3 Procedure for Data Collection

Data were collected primarily through interviews. These interviews used a semi-structured, interview style, which allows the researcher to adhere to a specific area of inquiry, but also retains the freedom to probe further into topics when needed (Miles & Huberman, 1994, p.16; Creswell, 2007, p.132). More specifically, this

method fits in the middle between a tight and loose protocol (Miles & Huberman, 1994).

These interviews may also be described as having a focused interview style, as represented by both the amount of time taken to conduct the interview as well as the specificity of the interview protocol (Yin, 2009, p.107). The focused interview approach was chosen for two reasons: First the research topic was identified prior to the interviews; and Second, the time commitments needed for an in-depth interview would be too great for most interview candidates (Yin, 2009, p.107). The interviews were audio recorded, and notes were taken. The interviews lasted approximately between 45 minutes and 1 hour.

The interview protocol can be found in Table 2 of Appendix B. As Table 2 shows, all the interview questions correspond to aspects of best practice found in the private sector literature on performance appraisal systems. The questions were designed to develop insight into the difference between private sector and INGO performance appraisals. To further gain insight to the functionality of the performance appraisal system being used, probes addressed accountability in all directions.

*Insert Table 2 here *

4.4 Procedure for Analysis of Results

The data collected were analyzed using primary coding and theoretical analysis (Miles & Huberman, 1994). When interview data were collected, a general reflective summary was created to help the researcher take a larger and more holistic viewpoint of the data (Miles & Huberman, 1994). This enabled the researcher to reflect on larger trends, which, in turn, informed subsequent, more specific analysis of the data (Creswell, 2007; Miles & Huberman, 1994). More specifically, the data collected from the interviews were then compared and contrasted with key points from the literature review to allow for further insight to the questions of how performance appraisal systems are conducted, in INGOs. This analytical approach allows for a much higher level of insight, it also contributes to a more accurate outcome as findings found from data collection will be directly tied to previous validated theories and concepts (Corbin & Strauss, 2008). In addition to permitting the emergent themes to surface (Corbin & Strauss, 2008; Miles & Huberman, 1994) it is through this process that the larger themes became more apparent to the researcher. It also aided in refining the interview protocol. Thus a combination of the data from organization, and reference to the theoretical framework allowed the general themes to emerge into more specific findings.

Coding the data was initially going to be done by using qualitative analysis software called NVIVO. However, this proved too costly for the author, so manual coding was employed instead. The coding was then conducted again by a second researcher to help increase the accuracy of the coding being done (Creswell, 2007). This double

step approach, while more time consuming, is a significantly safer and more reliable approach to analysis for a new researcher (Corbin and Strauss, 2008). The descriptive codes originally used were used to help create pattern coding that emerged from the cases, and the literature. This effectively allowed the study to use a blend of inductive and a deductive approaches to coding, which helped add both legitimacy as well as context to the themes being discovered (Miles & Huberman, 1994; Yin 2009).

This study also used member checking to help increase the level of accuracy in the data collected, as well as help ensure the conclusions found are also correct (Creswell, 2007). Member checks were conducted when the interview transcripts were complete, as well as when the initial findings were ready. This multi-step verification process not only allowed for accuracy in the findings, but also helped avert time inefficiencies; Namely, checking with respondents throughout the process lowered the risk for misinterpretations in the final report.

5. Results

5.1 Overview

Because of the small size of the organization, departmental identifiers were removed to preserve anonymity. Fortunately, due to the structure of the interview protocol, the majority of the relevant quotations were able to be used without compromising anonymity. The entire staff was interviewed for the purposes of the research, including 3 managers and 18 employees.

The results section is organized in a similar fashion to that of the literature review. Namely, it was organized according to various elements of a high performing (“best practices”) performance appraisal system. This approach was intended to facilitate the contrasting of results vs. theory.

5.2 Research Question #1 – Accountability

The first research question for this paper was:

RQ1 How, if at all, is the need for accountability in INGOs reflected in how they conduct appraisals?”

The literature presented a four-path approach to looking at how an INGO needs to be accountable (upward to donors, downward to beneficiaries, inward to employees and outward to peer NGOs in the sector).

When examining the results (see Table 3.1A,B), it became clear that interviewees acknowledged a focus on two of the four main forms of accountability.

Upward accountability (to donors) was a prominent issue for many of the interviewees; There was a strong relation between the work being completed and awareness of the need to ensure the donations received from donors were to be used effectively and efficiently. This was made apparent by one manager's comments.

“We are servicing the public, both the donors and the recipients. We are using other peoples money to do things, not our own. We need to change the way we do business, update it. We can't have 4 people taking up half a million in salary. We need to streamline, be more effective in the delivery of our programs. There is too much waste in the system, waste that donors [do not want to] pay for.” (M1)

Relatedly, interviewee's comments conveyed recognition of the need for board accountability as well.

“There was a core group of people [on the board] doing most of the work and some individuals who would sit back and not help. So one of the changes, is to bring in board terms [of reference (e.g. operating regulations)], and enforce them, [including having] board transitions.”
(M2)

On the other hand, inward accountability (to the staff) was acknowledged but reactions were less positive. It appears that management and staff are well aware that there is a need for individual accountability. But, apart from the occasional positive response (e.g. E4), there is generally perceived to be a lack of effective accountability towards employees. Problems with inward accountability included a sense of excessive micromanagement (Time Tyranny), which most likely stems from trying to ensure everyone works equally as hard, but as one employee points out, it can have the opposite effect.

“Watching us every minute of the day to make sure we are doing what we are supposed to because it’s on [working] hours, that’s not right. So if that is what that would do [as a result of a new performance appraisal system], we may as well all quit. It wont be [*organization*] anymore, it will be something else.” (E1)

The issues around micromanagement are also further compounded by a perception of distributive justice lacking in the organization.

“You know right now it is kind of a mystery as to why we get a pay increase and why we don’t. That part really is not clear, it’s almost like they just decide randomly to give them out or not. Well that is not true, they gave us all a big one awhile back, but not because we were doing well, it was that we were all being paid so poorly.” (E2)

This employee (as well as many others) later go on to mention issues surrounding the procedural justice perceived in the organization.

“I mean we are all doing the same job essentially, just different programs, why not be evaluated the same. It’s like any other company, like working at a store, all cashiers have the same job, their customers served can change dramatically, but it is the same job nonetheless.” (E2)

In addition, there was felt to be a lack of HR expertise, which was considered to underline the above problems.

“Part of the problem here is not having a qualified HR staff member. Do we need a full time staff member? No. But do we need a professional here, yes. If we are supposed to grow... we need a qualified individual to lead the changes to make it happen. This is where we are at right now,

we need someone qualified to fix the issues to increase productivity.”

(M3)

Finally, when discussing their performance appraisal system, the respondents had no comments about the other two directions of accountability (downward and outward).

5.3 Research Question #2 – Private Sector Best Practices

The second research question for this paper was:

RQ2 How are the best HR practices for performance appraisals (as advocated in the private sector management literature) actually being applied to performance appraisal systems in INGOs? And if/when they do deviate, why?

This question was drawn from the sections of the literature review, which focused on the best practices for performance appraisals used in private sector. In the NGO sector, research suggests that international NGOs face additional barriers to applying private/public sector human resource management “best practices” due to their unique context (O’Sullivan, 2011). The following subsections therefore serve to contrast the current best practices for performance appraisals in private sector with the results actually found from the primary data collection.

5.3.1 Assessment Criteria

As stated previously, the criteria used to assess performance are fundamental to the effectiveness of the appraisal (Boswell & Boudrea, 2000, Pettijohn et al., 2001a, 2001b). The assessment criteria are the foundation for the entire performance appraisal. Without proper criteria, all other aspects of the appraisal will be greatly diminished. One of the main problems seen in performance appraisals is the incorrect assessment of tasks and roles, which lead to either criterion deficiency or criterion contamination (Boswell & Boudrea, 2000). This problem can arise in regard to the balance between judgmental and developmental objectives. As mentioned in the literature review, research favours a hybrid (output and behavioral input) approach over an exclusively input or exclusively output approach (Boswell & Boudreau, 2000; Morris et al., 1991; Pettijohn et al., 2001a; Smith et al., 2000).

Results (see Table 3.2A,B,C) indicated that assessment criteria suffered from both criterion deficiency and criterion contamination issues. In regard to criterion deficiencies, there were three main categories that emerged. Interviewees noted that in their performance appraisals staff numbers were unrecognized, which in turn could change the outcome of their performance and therefore performance appraisal drastically, as one employee makes clear in their comments.

“You know, [the managing director] is so determined to work off job descriptions, and it’s just not right. We do our job descriptions worth and

then some... When I started here, we had 3 people in the department and now we have [three times as many] including me. And that is not shown in the job description. It is still the old one, sure we still have the same countries and a few others but I have more than twice the staff. I have the entire organization come knock on my door and ask me to do something, sometimes demand it. And then I can get a lower rating because I am busy doing what I am supposed to in person, but not on paper. How is that fair? [How to] make it work? Make it based off what I do, not what I was supposed to do 8 years ago. I don't know how, but just fix it so it makes sense to my job and can show me what I need to do."

(E1)

This inability to change assessment criteria became a pattern after another employee discusses issues with increased work demand due to having students to mentor during a period of time, yet their change in work would not be properly measured due to this issue revolving around assessment criteria, as seen in another employee's comment.

"The only downside [of the performance appraisal system] is that it is based off our work plans, but as you know, our work plans can change drastically when new opportunities arise... I had an opportunity to have students from Algonquin College come down and help with some of the production of our rollouts. And this was good, but it ate up a lot of extra

time that I couldn't focus on other things. So while it was completely worth it, and overall a beneficial action, if it happened around performance appraisal time, it could skew the results cause something may not have gotten done in that past week or two." (E4)

These issues of unrecognized work tasks and the inability to assess them properly comes from the organization's need to use the hard criteria determined by CIDA instead of using a hybrid approach that could measure crucial qualitative metrics. This issue became evident with multiple comments by employees when they noted the fact that qualitative results were being neglected, as seen in the following comment.

"In theory [using CIDA work plans for our performance objectives] is a great approach. There is [nothing] wrong with what we hand in to CIDA... but the length of them, and level of detail is... not needed. I guess with that said, maybe less emphasis on the CIDA work plans, and more emphasis on actual [individual] performance work plans would also make a big difference [for the utility of the performance appraisals]. No more needing to read between the lines and such." (E5)

Results also indicated that criterion contamination occurred due to external issues such as problematic project partners and natural disasters. This contextual problem can be a potentially significant issue when using hard criteria, and it was made ever

apparent by the following comment by an employee when discussing issues with their industry.

“The thing about our business is there are certain things that trump all else. We were about to start up in [country x], but all the best laid work plans, evaluations, procedures couldn’t do anything against the fact that the country was destroyed by an earthquake.” (M2)

External issues such as these are common in the international development field, but participants clearly viewed it as problematic that the organization’s adherence to exclusively hard CIDA criteria precluded any qualitative consideration of these extraneous issues.

5.3.2 Instrumentation

As mentioned in both the section above and in the literature review, a performance appraisal instrument must be able to serve both judgmental and developmental purposes (Boswell & Boudreau, 2000; Morris et al., 1991; Pettijohn et al., 2001a; Smith et al., 2000). While there are multiple ways in which either one of these purposes types may be measured and achieved, the current state of research supports the use of both Management By Objectives (MBO) to tap hard (objective) criteria, and Behaviourly Anchored Rating Scales (BARS). This hybrid approach allows the organization to gain important performance metrics, while also aiding in the personal and professional growth of the employees.

Results from the instrumentation questions (see Table 3.3A) indicated that the participants felt that communication of the instrumentation use and process was done effectively.

“We have an annual evaluation for all employees. Using their job description as a reference and using the [CIDA] annual work plan as a reference, they [(the employees)] essentially state (according to a certain number of questions) how they feel they performed over the course of that year. Their manager looks at that and adds their own comments... They then meet together to talk about it, and the employee has a chance to write on the document, their take and their feeling about what feedback they got from the manager.” (M2)

However, responses also indicated a strong emphasis on judgmental outcome data, to the exclusion of behavioral (developmental) data. For example, the outcome measures for the employees are closely aligned with the CIDA outcome objectives for the various projects (as per the project proposal). This meant an inconsistent approach across departments, which have different types of projects (and varying numbers of project objectives).

“One of the things we are looking at is how to have more of a common format for individual work plans [performance objectives] across the entire organization. They require these long very detailed [Result Based

Management] output/outcome results statements. The trouble with these documents is they tend to be too [output] activity oriented. So it is page after page of “this workshop will happen, this trip will happen, this report will be written”. As a useful document it’s too detailed, and too focused on the activity level that doesn’t really inform performance. [When translating this into our individual performance appraisal instruments] we need more of a common approach across the organization so that if you were looking at this department, and this department and this department, that they are all apples. You know that it is not different system [being used] in different departments.” (M2)

In addition, the unbalanced focus on outcome measures was accompanied by rigidity (i.e. it was not responsive to amendments needed to acknowledge additional behavioral demands, such as tripling of staff responsibilities).

“We [my colleague and myself] came up with the idea of having an annex [to my performance appraisal and job description] to reflect that we have increased to 8 other staff since I started. But my manager later denied this request.” (E1)

Further concerns about the developmental deficit of the performance appraisal are addressed in the coaching section.

5.3.3 Biases and Rater Training

It is generally accepted (with some mixed reviews) that ratee sex, age and gender affect performance outcome insignificantly (Bretz, Milkovich & Read, 1992). However, there are still biases that occur on an unconscious level due to human error (Bretz et al., 1992). Perceptual problems raise legality issues. While it is impossible to completely eliminate the chances of human error, with proper rater training there can be significant improvements in the effective use of the rating tools by significantly decreasing human error (Bretz et al., 1992). Therefore, it is essential for all organizations to properly train their raters in the tools they are using to assess employees.

The results indicated that most candidates were happy with how the appraisal process went. However, they also saw key flaws in how it was executed, which they felt warranted rater training. Two main themes emerged in relation to biases and rater training (see Table 3.4A,B). First, in terms of biases, several interviewees identified a tendency for a positive bias. One manager points out this bias well, and helps to indicate that this bias is due to poor rater training, as they are favouring actions that they perceive to elicit desirable social outcomes, while as the literature shows, these actions could leave the employee feeling less satisfied with the assessment and the ratee as well.

“[With] most of the [performance appraisals] I read, it is 95% positive... They are all: “We did this.” “This happened.” “We did that.” It is almost just a listing of great things that happened. There is nothing wrong with that, but rarely do I see any tough aspects come up. Like, “You are really struggling with this element, or you really seem to have a hard time moving ahead on this”. I think the thing there is that it is a small organization, and each department is its own group, and I think that every director wants the loyalty of their group, so there is a certain incentive to, not be too critical of anybody or anything. And you can’t really criticize anybody outside your group because it is not really your business. So there is a natural, in-born motive to not ask hard question or to not pursue [critical] things too far.” (M2)

Second, interviewees noted inconsistency among managers’ performance appraisal approaches. Accordingly, respondents pointed to a need for rater training to improve the reliability and utility of the performance appraisal.

“I think [the performance appraisal instrument and process] is a very effective tool for identifying my strengths and weaknesses. But I also think that it can be a very poor tool for it as well. It really comes down to how honest both parties are in the process, and how well trained you are at administering them as well. [In response to the question of what is needed to improve the performance appraisal process] I would say proper training

for everyone. Better execution of the system will help both managers and employees fill out these forms better, causing a lot less stress and a lot more functionality. I... see them being easily rushed over, and just kind of checking each one off and getting it done ASAP, but I also see it being done slowly and methodically. One approach will give great results, and other results will give poor results. While [my supervisor] has a little bit of training in the area, he is just a natural leader and coach for the team, and is always there for the extra mile of help. If we didn't have [our manager], [the appraisal discussions] would be far less beneficial. And like I mentioned about other managers previously, I can see their employee's appraisals being rather useless. I don't think there is proper training on how to coach and evaluate and how to correct behaviours. So it comes down to the natural ability of the manager, which in my case makes me lucky, but others may not be." (E5)

5.3.4 Coaching

Coaching is an essential element of employee development in the performance assessment process. Through proper coaching, employees should have a better understanding not only of their actions and the outcomes but also of future steps needed to improve (Bretz et al., 1992). As reviewed from the literature, an increase in pay or a promotion will leave the employee happy, but they may not be satisfied with the appraisal discussion if it was done in a proper or conveyed in an ineffective manner (Bretz et al., 1992). To be effective, performance feedback in a coaching

session should be supportive, emphasize improvement, and discuss advancement. If done in this manner, employees will have a higher level of satisfaction with the process (Bretz et al., 1992).

As Table 3.5A indicates, interviewees generally reacted positively to their coaching experience and indeed noted it to be one of the key beneficial processes the organization uses. Most of the remarks indicated not only that there was supportive performance feedback present in the formal evaluation process, but that they benefited from frequent day-to-day coaching interactions, as seen by this employee's remarks.

“We fill out a standard [self appraisal] form about how we feel we have done, and then we go have that discussion. At which point [Managing Director] will make notes on anything, critique if you will. I think it is a pretty simple process that way. The [supervisor's feedback on our self-appraisal in the form of] notes really are stopping people from grossly exaggerating what they are doing and such. So yeah we do this, then we talk about how I can improve and look at my work plan and go from there. It's very much a give and take process, rather than just a verdict being passed down kind of thing. It was nice to sit down and talk about things formally, rather than just the constant... less formal way of a quick pop in meeting.” (E4)

When criticisms were raised about coaching, it was more about the performance appraisal instrument causing the coaching process to be too detailed (too results-based rather than developmental based) and too inconsistent across managers. And this can be seen by one manager's comments about the appraisal process being used.

“The non-development [(i.e. the judgmental part of the formal performance appraisal)] that was easy to sort of present and talk about, because the development [part of the discussion] was so far down in the weeds, that, I think that was the problem. I think [the annual performance appraisals don't] go deep enough on how an employee can be supported to do more and be more effective. And because it is annual, it is less than an informative process it becomes more of a summative process. So it comes down to, “Did you do it or not?” Perhaps a monthly or perhaps a six month [(bi-annual)] process may lend itself to more coaching and looking at what can be brought in, professional training. And [that could more effectively] identify weakness, and [enable the manager to] say you need to go take a course in something, or we all need to take a course in something.” (M2)

5.3.5 Sources of Appraisal

As the literature review discussed, a high performance performance-appraisal system strives to enhance development outcomes by using as many perspectives as possible (Antonioni & Park, 2001). In most cases this would mean a self, peer and a supervisory evaluation (Antonioni & Park, 2001). Employees who are also managers may also have their employees perform an evaluation on them (Antonioni & Park, 2001).

Results regarding sources of appraisal were somewhat consistent with the literature (see Table 3.6A). All interviewees indicated that two sources of appraisal were used (self and supervisor), as per the literature review's recommendations. The following comment from an employee represents the view of every other employee interviewed.

“Once a year I will fill out a self-evaluation form. And then I hand it into [Managing Director] where he will review it, and then we meet about it.”
(E6)

However, upward and peer sources of appraisal were not reported. The exclusion of a fully 360-degree approach may have been due to the small size of the organization, which would have risked compromising anonymity.

5.3.6 Frequency

Frequency is the final important element of an effective performance appraisal system. It can play a strong role in contributing to perceptions of validity and reliability of the process being used (Pettijohn et al., 2001a). It is most common to see a single yearly evaluation being conducted by an organization to help minimize time costs. However this may in fact be a detriment to both the organization and the process (Pettijohn et al., 2001a). Often, managers can only remember the past 3-4 months accurately, and will either ignore the remaining 8-9 months or use poor judgment on longer time frames (Pettijohn et al., 2001a). Having a more frequent cycle for formal feedback will allow for more accurate performance appraisals, while also increasing employee's acceptance of the process (Pettijohn et al., 2001a). More specifically, with bi-annual formal performance appraisal sessions, management will have the ability to reduce the chance of bad habits forming while also giving a more accurate appraisal; this, in turn will aid the organization and the employee more (Pettijohn et al., 2001a).

Table 3.7A,B demonstrates that the frequency of formal appraisals occurs once a year in the participating INGO, which is consistent with the common approach noted in the literature review. However, the timing of this formal annual appraisal was regarded as onerously overlapping with key (and busy) business process times for the organization (e.g. annual reporting to CIDA).

“To really make use of [the performance appraisal feedback process] though, we’ve got to have more of a day to day, if not... [at least a] more regular approach [to] performance management, as opposed to once a year. We are still looking at it at the end of the year. And by that time, of course it is too late. Our annual budget process is in February, and [we need] to put forward a budget... at the March Board meeting to have it approved by the beginning of the new fiscal year, April 1st. [So] if budget and employee compensation was linked to job performance, then you would want to have those evaluations done, say in December or January, so that would factor into the budget process. And instead we tend to do it, we tend to get everything else out of the way, and in April we finally get around to it.” (M2)

Fortunately, any shortcuts taken in the performance appraisal discussion (due to the poor timing) appeared to be rectified by informal coaching from management, which occurred more frequently. This informal coaching was felt to offset the drawbacks of a single annual formal evaluation being executed at a busy time of the year. This is expressed clearly in the following employee’s comment about the frequency of evaluation.

“Formally they [the performance appraisals] are done annually, but in actuality it is almost a constant evaluation or conversation about performance... If you are told everything is going well... [then] sudden [concerns about performance] should be gradual.” (E5)

6. Discussion

6.1 Summary and Theoretical Implications

There are four key findings that emerge from this study. First, a successful performance appraisal system is greater than the sum of its parts. The inclusion of best practice elements for a performance appraisal system are no guarantee of successful execution. Tacit HR expertise is needed to increase the probability that those elements work harmoniously together. This study demonstrated that even aspects of best practices that require considerably less HR expertise (e.g., scheduling the performance reviews) may be executed poorly without dedicated managerial evaluation of the performance appraisal system. Second, when accountability to donors leads an INGO to unquestioningly apply donors' project-level performance criteria to individual-level performance appraisals, it leaves the performance appraisal system vulnerable to both criterion deficiency and criterion contamination, which represents a liability. Third, introduction of a private sector tool to a nonprofit organizational context represents a significant culture change, and requires a systematic organizational change intervention. Finally, one unexpected advantage of the INGO's culture of a shared ethos for human resource development is that it is conducive to frequent informal coaching. The theoretical implications of each of these findings will be discussed in more detail below.

6.1.1: Finding #1: A Successful Performance Appraisal System is Greater than the Sum of its Parts

The literature (e.g. O'Sullivan, 2010; Ramia, 2003; Labelle, 2007; Sachs, 2007; Cavill & Sohail, 2007) has repeatedly emphasized the need for increased accountability in NGOs. Although respondents acknowledged general upward accountability, they had considerable concerns about how inward accountability was being attempted in the organization. Specifically, one of the key findings of this study is that many elements of "best practice" performance appraisals may be present, but the overall performance appraisal system may still fail to fully achieve its goals if organizational dynamics are given less consideration. Several examples of this were evident:

First, output criteria were favoured over input criteria, with insufficient consideration given to the well-documented fact that the latter are critical to procedural justice (Smith et al., 2000). Second, there was also clearly a sense of micromanagement, and rigidity in terms of the manner in which the appraisals were conducted, despite the fact that micromanagement is another known pitfall commonly addressed by the performance appraisal literature (e.g. Pettijohn et al., 2001). Although coaching can be particularly useful for organizations like INGOs in which a pay increase or promotion is not possible and/or warranted (Bretz et al., 1992), it needs to be conducted in a supportive manner that encourages improvement and motivates by offering rewards of some kind (even if not monetary), rather than in a micromanaging manner that risks impairing distributive justice. Third, participants acknowledged that the ethical ethos of responsibility for

human resource development that they shared with their like-minded appraisers may have led their appraisers to be less critical in the performance appraisals (as evidenced by the comments in Table 3.4A). Yet they also acknowledged a lack of rater consistency in this regard. Overall, they attributed raters' inconsistency to a lack of proper rater training, and this appeared to affect their acceptance of the performance appraisal system. This too, supports prior research on the importance of rater training (Karriker & Williams, 2009; Heneman, 2007), and extends that research to the INGO context.

Fourth, there were aspects of the performance appraisal system that were fully within the control and expertise of the INGO, such as the frequency with which (and the timing when) they hold formal appraisals, but the INGO did not fully avail itself of these. A satisficing approach was observed, with formal appraisals being conducted only annually, and scheduled at an onerous time of year. This annual-only approach is consistent with the private sector's findings that most firms do the same (e.g., Pettijohn et al., 2001a). This study thereby extends that research to the INGO sector. In addition, this finding contributes to the NGO management literature, which has sparsely examined the frequency of appraisals (Lewis, 2007), or any other elements of a high performance appraisal system. However, it is also consistent with the literature's observations that the time-related cost-savings of annual-only appraisals come at the expense of lower utility of the appraisal tool, and potentially lower credibility of the tool for employees, thereby risking weakened acceptance (e.g., Pettijohn et al., 2001a). Clearly, as the management literature

suggests, a bi-annual formal appraisal is preferred because it reduces errors due to memory loss (e.g., Pettijohn et al., 2001a).

All of these observations about how the elements of the performance appraisal system were executed affirm the findings of existing private sector research on organizational justice (Karriker & Williams, 2009; Heneman, 2007) and extend it to the INGO sector. Namely, when issues of both procedural and distributive justice are overlooked in a performance appraisal system, it can greatly diminish the level of employee acceptance, thereby obstructing the effectiveness of the process.

However, whereas many organizations in the private and public sectors may also fall short in addressing organizational dynamics when implementing a high performance appraisal system, this type of shortfall may be more probable in INGOs where HR is often treated as a luxury. In INGOs, activities such as training are costed as overhead, and donor regulations around overhead versus program expenses are rather strict (O'Sullivan, 2007). Thus, the INGO context is one in which acquiring the tacit HR knowledge (i.e., to implement a performance appraisal system using proper organizational change approach) can be financially difficult. This broader issue -- of a lack of internal tacit HR knowledge within the organization-- appears to have created a performance appraisal environment that, on the surface, checks off the appropriate performance appraisal steps, but in reality is impaired in its ability to consistently and accurately assess the performance of the employees in an INGO

6.1.2: Finding #2: Uncritical Accountability to Donors Raises Potential Liability

An emergent finding concerns how the INGO's dependence on CIDA's financial resources seems to have influenced the INGO's decision to use the program-level performance metrics as individual level output criteria. There appears to have been little initiative on the part of the INGO to negotiate with CIDA for more suitable and balanced (more development, input-based) performance metrics. Negotiation is required because, fundamentally, there is a paradoxical conflict between the two organizations' (CIDA and the INGO) objectives. They share the same long-term goal, but diverge on the interim objectives: CIDA wants to see efficiency, but there are additional things that an INGO needs to do to be effective which are not reflected in the hard outcome criteria. Neglecting these softer, input criteria aggravates the sense of organizational injustice with the performance appraisal system (e.g., Heneman, 2007), and, in that sense, weakens it. It also leaves an INGO vulnerable to accusations of liability when adherence to CIDA's hard criteria is prioritized over recognition of valuable input metrics or other confounding factors:

“They [CIDA] also sometimes require a work plan that is far too detailed in certain areas and not nearly detailed enough in others. So if we were to solely guide our performance based off of these work plans, you would either have poor performing workers, or a poor performing company.”

(E5)

Clearly, the application of a donor agency's cookie-cutter, hard criteria approach to implementing HR procedures and policies does not do justice to either the government or the INGOs, nor, by extension, to the beneficiaries either. A lack of hybrid criteria may be problematic even in the private sector, but it is particularly problematic in the international development sector, which is routinely subject to so much extraneous volatility (Pettijohn et al., 2001a, 2001b; Boswell & Boudreau, 2000; Smith et al., 2000; Li & Bagger, 2008; Pettijohn and Kent 2001; Smith, Harrington and Houghton, 2000; Leandri, 2001; Bachrach, Bendoly & Podsakoff, 2001).

6.1.3: Finding #3: Introduction of a Private Sector Tool to a Nonprofit Organization Potentially Represents a Culture Change

The findings of this study suggest that the introduction of a formal performance appraisal system to an NGO must be recognized as an organizational change initiative. A formal performance appraisal system is the artifact of a private sector culture. As such, it represents a shift away from an INGOs' historically un-judgmental culture that was founded in the shared compassion for human resource development. As Dougherty (1992) noted, people may agree on what needs to be done, but it may be quite another issue to agree on how to properly do it; disagreement is often founded in functional differences. Therefore, HR expertise is needed to orchestrate this change initiative effectively, and carefully thread the new

tool into the culture of the INGO. However, given funding constraints, the question of how to acquire that HR expertise is less than straightforward.

6.1.4: Finding #4: The INGOs' Cultural Value for Human Resource Development is an Asset

On the other hand, the shared sense of community and ethos for human resource development that exists among the INGO's employees appears to have lent itself to more frequent informal coaching. The prevalence of frequent, informal coaching provided thus compensates somewhat for the weaknesses with the formal process, and clearly helped to improve employees' experience of the performance appraisal. This finding contributes to the NGO management literature, which has sparsely examined the how performance appraisals are conducted in INGOs (Lewis, 2007). Nevertheless, because managers differed considerably even in their approach to informal coaching (see Table 3.5A), the major themes for improvement identified earlier remain in need of closer attention.

6.2 Practical Implications

The practical implications from this paper can be directed at two target audiences; 1) Changes CIDA needs to make to help their NGO partners improve performance, and 2) The changes the INGOs need to make to improve their performance appraisal function.

First, CIDA needs to begin by looking at the reporting requirements they are using for assessment of effectiveness in NGOs and especially development INGOs as it has been made clear that objective measures do not capture the performance of the organizations properly or as completely as they intend to. By broadening their assessment criteria and reporting requirements of the organizations they deal with, they can essentially eliminate the counterproductive exclusive focus on assessment objectives that they have institutionally created. Specifically, they must rectify the unbalanced performance appraisal criteria currently being encouraged in the INGOs, in order to better allow these organizations to properly develop their employees, and their organization, so that they may efficiently achieve performance objectives.

Second, to persuade CIDA to do this, Canadian INGOs should collaborate to encourage CIDA to modify their reporting system. They should strive to enable it to be more dynamic (to allow for contextual differences), and better reflective of contingencies being experienced by these organizations.

Third, INGOs need to recognize the need for increased HR expertise amongst their staff. This will not only eliminate many of the business process slow downs created by having poorly formulated policies and procedures, but will also increase productivity in the long run by having employees who possess a greater sense of justice and who are properly developed. If excessive dependence on CIDA funding is a constraint, the INGOs need to dilute their resource dependency on CIDA (Pfeffer,

& Cohen, 1984; Pfeffer, & Salancik, 1978) by seeking other funding sources to help them acquire critical, tacit HR expertise.

6.3 Limitations and Directions for Future Research

6.3.1 Limitations

The main limitation of this research paper is that it was conducted with a small sample size (n=21) and a single organization. However, this sample represented the entire size of the INGO. Moreover, it should be noted that the INGO sector primarily consists of small organizations. Thus, conducting this type of research in any one INGO would normally involve smaller numbers than similar research conducted in multinationals within the private sector. Furthermore, although two to three organizations would have permitted cross case comparisons, the focus on this single organization permitted more in-depth analysis.

6.3.2 Directions for Future Research

In terms of future research directions, the first suggestion is for future studies to rectify the limitations noted above. Namely, a similar study should be completed on one or two more education-focused development INGOs in Canada to see if similar results are to be found. If so, further grounding in some of the ideas presented in this paper can then be explored on a large scale.

Secondly, future research might compare the results in this INGO with INGOs that have a higher level of HR expertise. Finding an INGO which employs an HR professional would permit helpful comparison, and help reveal if the tacit HR knowledge issues brought up in this paper are critical for the sector as a whole, or are unique situational factors needed by this specific organization.

A third avenue for future research would be to explore the utility of having the INGOs rely on the umbrella organization of all Canadian INGOS, the CCIC (Canadian Centre for International Cooperation) as an advocacy group to CIDA. The study could explore the utility of this umbrella organization in eliciting change from CIDA (e.g. to in regard to developing a more robust set of performance objectives).

Finally, given that the trend in international development funding in Canada is in the direction of private sector-INGO alliances (Rangan, Samii, & Van Wassenhove, 2006; Ruffin, & Rivera-Santos, 2012), the need for recognition of organizational development practices to permit improved performance management has never been greater. Therefore, future research could compare diverse INGO-private sector alliances to explore whether INGOs allied with private sector organizations may benefit from the private sector's HR expertise.

7 References

- Aiken, C., & Keller, S. (2009). The irrational side of change management. *The McKinsey Quarterly*, 2, 102-109.
- Akhtar, S., Ding, D. Z., & Ge, G. L. (2008). Strategic HRM Practices and their Impact on Company Performance in Chinese Enterprises. *Human Resource Management*, 47(1), 15-32.
- Antonioni, D., & Park, H. (2001). The relationship between rater affect and three sources of 360-degree feedback ratings. *Journal of Management*, 27(1), 479-495.
- Bach, J., & Stark, D. (2002). Innovation Ambiguities: NGOs' Use of Interactive Technology in Eastern Europe. *Studies in Comparative International Development*, 37(2), 3-23.
- Bachrach, D., Bendoly, E., & Podsakoff, P. (2001). "Attributions of the 'causes' of Group Performance as an Alternative Explanation of the Relationship between Organizational Citizenship Behavior and Organizational Performance", *Journal of Applied Psychology*, 86(6), 1285-1293.
- Becker, B. E., & Huselid, M. A. (1999). Overview: Strategic human resource management in five leading firms. *Human Resource Management*, 38(4), 287-301.
- Beem, M. J. (2001). Fundraising in the balance: An analysis of job performance, appraisals and rewards. *International Journal of Nonprofit and Voluntary Sector Marketing*, 6(2), 164-171.
- Beer, M. and Nohria, N. (2000) *Breaking the Code of Change* (Boston: Harvard Business School Press).
- Behery, M. H., & Paton, R. A. (2008). Performance Appraisal-Cultural fit and organizational Outcomes within the U.A.E. *Journal of American Academy of Business*, 13(1), 166-176.
- Binning, J., & Barrett, G. (1989). Validity of Personnel Decisions: a Conceptual Analysis of the Inferential and Evidential Bases. *Journal of Applied Psychology*, 74(3), 478-494.
- Boswell, W. R., & Boudreau, J. W. (2000). Employee Satisfaction with Performance Appraisals and Appraisers: The Role of Perceived Appraisal Use. *Human Resource Development Quarterly*, 11(3), 283-299.
- Bretz, R. D., Milkovich, G. T., & Read, W. (1992). The Current State of Performance Appraisal Research and Practice: Concerns, Direction, and Implications. *Journal of Management*, 18(2), 321-352.
- By, R. (2005) Organizational change management: a critical review, *Journal of Change Management*, 5(4), 369-380.

- Cavill, S., & Sohail, M. (2007). Increasing strategic accountability: a framework for international NGOs. *Development in Practice*, 17(2), 231-248.
- CIDA (2006) Table G: List of CIDA disbursements made to voluntary sector and international voluntary sector by delivery channels for 2004-2005. CIDA's statistical report on official development assistance, fiscal year 2004-2005 (pp.25-27). Report prepared by the Statistical Analysis & Reporting Section, Information Architecture & Products, Information Management & Technology Branch, Canadian International Development Agency.
- Corbin, J. M., & Strauss, A. L. (2008). *Basics of qualitative research: techniques and procedures for developing grounded theory* (3rd ed.). Los Angeles, Calif.: SAGE Publications.
- Crawford, L., & Nahmias, A. H. (2010). Competencies for managing change. *International Journal of Project Management*, 28(1), 205-412.
- Creswell, J. W. (2007). *Qualitative inquiry & research design: choosing among five approaches* (2nd ed.). Thousand Oaks: SAGE Publications.
- DeCarlo, T. E., & Leigh, T. W. (1996). Impact of Salesperson Attraction on Sales Managers' Attributions and Feedback. *Journal of Marketing*, 60(2), 47-66.
- Decotiis, T., & Petit, A. (1978). The Performance Appraisal Process: A Model and Some Testable Propositions. *The Academy of Management Review*, 3(3), 635-646.
- DeVries, D.L., Morrison, A.M., Shullman, S.L. and Gerlach, M.L. (1981), Performance Appraisal On The Line, Center for Creative Leadership, Greensboro, NC.
- Dougherty, D. (1992). Interpretive Barriers to Successful Product Innovation in Large Firms. *Organizational Science*, 3(2), 179-202.
- Drexler, JR., J. A., Beehr, T. A., & Stetz, T. A. (2001). Peer Appraisals: Differentiation of Individual Performance on Group Tasks. *Human Resource Management*, 40(4), 333-345.
- Fisher, L. (1994) Total Quality: hit or myth?. *Accountancy*, 22(4), 50-51.
- Ford, D. K. (2004). Development of a performance appraisal training program for the Rehabilitation Institute of Chicago. *Journal of European Industrial Training*, 28(7), 550-563.
- Fyvie, C., & Ager, A. (1999). NGOs and Innovation: Organizational Characteristics and Constraints in Development Assistance Work in The Gambia. *World Development*, 27(8), 1383-1395.
- Globerson, S. (1984). Issues in developing a performance criteria system for an organization. *International Journal of Production Research*, 23(4), 639-646.
- Grawitch, M. J., Granda, S. E., & Barber, L. K. (2008). Do Prospective Workday Appraisals Influence End-of-Workday Affect and Self-Monitored Performance?. *Journal of Occupational Health Psychology*, 13(4), 331-344.

- Guion, R., & Highhouse, S. (2004). *Essentials of Personnel Assessment and Selection*. New York: Psychology Press.
- Hempel, P. S. (2001). Differences between Chinese and Western managerial views of performance. *Personnel Review*, 30(2), 203-226.
- Heneman, H. G. (2007). *Staffing organizations: recruitment and selection in Canada* (Canadian ed.). Toronto: McGraw-Hill Ryerson.
- Heslin, P. A., Latham, G. P., & VandeWalle, D. (2005). The Effect of Implicit Person Theory on Performance Appraisals. *Journal of Applied Psychology*, 90(5), 842-856.
- Jennings, T., Palmer, J. K., & Thomas, A. (2004). Effects of Performance Context on Processing Speed and Performance Ratings. *Journal of Business and Psychology*, 18(4), 453-463.
- Karriker, J. H., & Williams, M. L. (2009). Organizational Justice and Organizational Citizenship Behavior: A Mediated Multifoci Model. *Journal of Management*, 35(1), 112-135.
- Knowles, R. (1999) Continuous change, *CA Magazine*, 132(4), 43-44.
- Kotter, J. (1999) Leading change: eight steps to transformation, in: J. Conger, G.Spreitzer and E. Lawler (eds) *The Leader's Change Handbook* (San Francisco: Jossey-Bass).
- Labelle, H. (2007). How do we break the corruption cycle? *The Globe and Mail*.
- Larson, JR., J. R. (1984). The performance feedback process: a preliminary model. *Organizational Behavior and Human Performance*, 33(1), 42-76.
- Leandri, S. (2001). "Measures that matter: How to fine-tune your performance measures". *Journal of Quality and Participation*, 24(1), 39-41.
- Lewis, D. (2007). *The management of non-governmental development organizations* (2nd ed.). London: Routledge.
- Li, A., & Bagger, J. (2008). Role ambiguity and self-efficacy: The moderating effects of goal orientation and procedural justice. *Journal of Vocational Behavior*, 73(3), 368-375.
- Luxmore, S. R., & Hull, C. E. (2011). Innovation and NGOs: A Framework of Interaction. *Entrepreneurship and Innovation*, 12(1), 17-27.
- Mastracci, S. H., Newman, M. A., & Guy, M. E. (2005). Appraising Emotion Work: Determining Whether Emotional Labor is Valued in Government Jobs. *American Review of Public Administration*, 36(2), 123-138.
- Milliman, J., Nason, S., Zhu, C., & Cieri, H. D. (2002). An Exploratory Assessment of the Purposes of Performance Appraisals in North and Central America and The Pacific Rim. *Human Resource Management*, 41(1), 87-102.
- Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: an expanded sourcebook* (2nd ed.). Thousand Oaks: SAGE Publications.

- Morris, M. H., Davis, D. L., Allen, J. W., Avila, R. A., & Chapman, J. (1991). Assessing the Relationships Among Performance Measures, Managerial Practices, and Satisfaction When Evaluating the Salesforce: A Replication and Extension. *The Journal of Personal Selling & Sales Management*, 11(3), 25-35.
- Narcisse, S., & Harcourt, M. (2008). Employee fairness perceptions of performance appraisal: a Saint Lucian case study. *The International Journal of Human Resource Management*, 19(6), 1152-1169.
- Neely, A., Mills, J., Platts, K., Richards, H., Gregory, M., Bourne, M., et al. (2000). Performance measurement system design: developing and testing a process-based approach. *International Journal of Operations & Production Management*, 20(10), 1119-1145.
- Noe, R. A. (2005). *Fundamentals of human resource management* (Canadian ed.). Toronto: McGraw-Hill Ryerson.
- Osberg, S. 2006. Skoll Foundation overview.
[http://www.skollfoundation.org/media/published_works/sosberg/Stanford%20June%2006%20%20Skoll%20Foundation%20Overview.ppt#263,8,Slide 8](http://www.skollfoundation.org/media/published_works/sosberg/Stanford%20June%2006%20%20Skoll%20Foundation%20Overview.ppt#263,8,Slide%208) (accessed 12 January 2009)
- O'Sullivan, S. L. (2010). International human resource management challenges in Canadian development INGOs. *European Management Journal*, 28(6), 421-440.
- O'Sullivan, S. L. (2011, Jan. - Feb.). A Look at Management Research in the Aid Industry: Are Humans the Forgotten Resource in International Development NGOs?. *Thunderbird International Business Review*, 53(1), 105-107.
- Pekrum, R., Goetz, T., Daniels, L. M., Stupnisky, R. H., & Perry, R. P. (2010). Boredom in Achievement Settings: Exploring Control-Value Antecedents and Performance Outcomes of a Neglected Emotion. *Journal of Educational Psychology*, 102(3), 531-549.
- Pettijohn, C. E., Pettijohn, L. S., & Taylor, A. J. (2000). Research Note: An Exploratory Analysis of Salesperson Perceptions of the Criteria Used in Performance Appraisals, Job Satisfaction, and Organizational Commitment. *Journal of Personal Selling & Sales Management*, 20(2), 77-80.
- Pettijohn, C. E., Pettijohn, L. S., & d'Amico, M. (2001). Characteristics of Performance Appraisals and Their Impact on Sales Force Satisfaction. *Human Resource Development Quarterly*, 12(2), 127-146.
- Pettijohn, C. E., Pettijohn, L. S., Taylor, A. J., & Keillor, B. D. (2001). Are Performance Appraisals a Bureaucratic Exercise or Can They Be Used to Enhance Sales-Force Satisfaction and Commitment?. *Psychology & Marketing*, 18(4), 337-364.

- Pettijohn, L. S., Parker, R. S., Pettijohn, C. E., & Kent, J. L. (2001). Performance appraisals: usage, criteria and observations. *Journal of Management Development, 20*(9), 754-771.
- Pfeffer, J., & Cohen, Y. (1984). Determinants of Internal Labor Markets in Organizations. *Administrative Science Quarterly, 29*(4), 550-572.
- Pfeffer, J., & Salancik, G. (2003). *The external control of organizations: A resource dependence perspective*. Stanford, Calif.: Stanford Business Books.
- Ramia, G. (2003). INGOs and Strategic Management. *Global Social Policy, 3*(1), 79-101.
- Rangan, S., Samii, R., & Van Wassenhove, L. (2006). Constructive Partnerships: When Alliances between Private Firms and Public Actors can Enable Creative Strategies. *Academy of Management Review, 31*(3), 738-751.
- Roberts, S., Jones, J., & Frohling, O. (2005). NGOs and the globalization of managerialism: A research framework. *World Development, 33*(11), 1845-1864.
- Ruffin, C., & Rivera-Santos, M. (2012). Between Commonweal and Competition Understanding the Governance of Public-Private Partnerships. *Journal of Management, 38*(5), 1634-1654.
- Salgado, J. (2001). Some Landmarks of 100 years of scientific personnel selection at the beginning of the new century. *International Journal of Selection and Assessment, 9*(1-2), 3-9.
- Schoorman, F. D., & Mayer, R. C. (2008). The Value of Common Perspectives in Self-Reported Appraisals: You Get What You Ask For. *Organizational Research Methods, 11*(1), 148-159.
- Shore, T. H., & Tashchian, A. (2002). Accountability Forces in Performance Appraisal: Effects of Self-Appraisal Information. *Journal of Business and Psychology, 17*(2), 261-274.
- Smith, A. D., & Rupp, W. T. (2003). Knowledge workers: exploring the link among performance rating, pay and motivational aspects. *Journal of Knowledge Management, 7*(1), 107-124.
- Smith, A. D., & Rupp, W. T. (2004). Knowledge workers' perceptions of performance ratings. *The Journal of Workplace Learning, 16*(3), 146-166.
- Smith, W. J., Harrington, K. V., & Houghton, J. D. (2000). Predictors of Performance Appraisal Discomfort: A Preliminary Examination. *Public Personnel Management, 29*(1), 21-32.
- Trochim, W. M., & Donnelly, J. P. (2008). *Research methods knowledge base* (3rd ed.). Mason, Ohio: Atomic Dog/Cengage Learning.

- Tziner, A., & Kopelman, R. E. (2002). Is there a Preferred Performance Rating Format? A Non-psychometric Perspective. *Applied Psychology: International Review*, 51(3), 479-503.
- Vyas, U. (2010). Japan's international NGOs: a small but growing presence in Japan-China relations. *Japan Forum*, 22(3-4), 467-490.
- Vales, E. (2007) Employees CAN make a difference! Involving employees in change at Allstate Insurance, *Organizational Change Journal*, 25(4), 27-31.
- Varma, A., Pichler, S., & Srinivas, E. S. (2005). The role of interpersonal affect in performance appraisal: evidence from two samples - the US and India. *The International Journal of Human Resource Management*, 16(11), 2029-2044.
- Walinga, J. (2008). Toward a Theory of Change Readiness: The Roles of Appraisal, Focus, and Perceived Control. *The Journal of Applied Behavioral Science*, 44(3), 315-347.
- Walsh, K., & Fisher, D. (2005). Action inquiry and performance appraisals: Tools for organizational learning and development. *The Learning Organization*, 12(1), 26-41.
- Whelan-Berry, K. S., & Somerville, K. A. (2010). Linking Change Drivers and the Organizational Change Process: a review and Synthesis. *Journal of Change Management*, 10(2), 175-193.
- Whiting, S. W., Podsakoff, P. M., & Pierce, J. R. (2008). Effects of Task Performance, Helping, Voice, and Organizational Loyalty on Performance Appraisal Ratings. *Journal of Applied Psychology*, 93(1), 125-139.
- Wiese, D. S., & Buckley, M. R. (1998). The evolution of the performance appraisal process. *Journal of Management History*, 4(3), 233-249.
- Wisner, J. D., & Fawcett, S. E. (1991). Linking Firm Strategy to Operating Decisions Through Performance Measurement. *Production and Inventory Management Journal*, 32(3), 5-11.
- Wong, K. F., & Kwong, J. Y. (2007). Effects of Rater Goals on Rating Patterns: Evidence from an Experimental Field Study. *Journal of Applied Psychology*, 92(2), 577-585.
- Wood, R. E., & Marshall, V. (2008). Accuracy and effectiveness in appraisal outcomes: the influence of self-efficacy, personal factors and organizational variables. *Human Resource Management*, 18(3), 295-313.
- Wright, P., Gardner, T., Moynihan, L., & Allen, M. (2005). The Relationship Between HR Practices and Firm Performance: Examining Causal Order. *Personnel Psychology*, 58, 409-446.
- Wright, R. P. (2002). Perceptual Dimensions of Performance Management Systems in the Eyes of Different Sample Categories. *International Journal of Management*, 19(2), 184-193.

Yin, R. K. (2009). *Case Study Research: Design and Methods* (Fourth ed.). Washington: SAGE.

Youngcourt, S. S., Leiva, P. I., & Jones, R. G. (2007). Perceived Purposed of Performance Appraisal: Correlates of Individual-and Position-Focused Purposed on Attitudinal Outcomes. *Human Resource Development Quarterly*, 18(3), 315-343.

8 Appendix A

Table 1. History of Appraisal Tools (Adapted from (Wiese & Buckley, 1998))

Title for Instrument	Description	Accountability	Judgmental and Developmental Coaching Purpose	Assessment Criteria	Overall Validity/Reliability
Global Rating/ Global Essay 1813	In 1813, an Army General began using a global rating system where he would use descriptions of men, such as; "a good-natured man" or "a knave despised by all". More Recent terminology used would be "outstanding", "satisfactory", "needs improvement".	May aid in accountability, but only in the fact that a rating system is being used. Does not offer higher accountability than any other system	Almost completely judgmental in nature, and will serve little to no developmental purposes	Uses a mixture of hard and soft criteria to make assessments. However, due to the subjective nature of the method, neither are used effectively to properly aid in job performance	Validity and reliability will be extremely poor as it is extremely subjective with little theoretical grounding
Man-to-Man Ranking 1914	Used by the army to rank officers, five scales were used: physical qualities; intelligence; leadership; personal qualities; and general value to service. The rater would then choose between 12-25 officers of the same rank as the officer being rated. Using the five scales, the rater will compare the ratee to these "benchmarks"	Will increase accountability slightly, but due to lack of job specific criteria, little accountability will be realized	Mostly developmental in nature, but due to the lack of consistency upon multiple raters, and high personal bias, will add little developmental capacity	Uses completely soft input criteria, but lacks jobs specific criteria	While reliability could be moderate to high, lack of job specific criteria limits validity drastically
Judgmental Rank Order 1920s	Raters provide an overall evaluation of performance by checking a box, which places each ratee in a certain percentage of all rates (top 25%, top 50%, bottom 50%, bottom 25%). A rater may also list each employee's name in order of effectiveness on	Accountability is higher than previous tools, but due to lack of job reference, the accountability	Minor judgmental abilities of this instrument due to the ranking, but mainly serves as a developmental	Depends on the criteria being used. But the focus is primarily on soft inputs	Reliability will be low, especially with multiple raters. Validity will be low due to lack of job reference

<p>Graphic/Trait Rating Scale 1920s-1930s</p>	<p>The rater indicates on a numerical scale the degree to which the ratee possesses certain personality traits. Performance dimensions are usually ill-defined and difficult to measure (leadership, initiative, creativity, etc). This is the first time we see a quantifiable approach to evaluations, instead of strictly qualitative like previous tools.</p>	<p>The non-job specific nature of this tool limits almost all accountability of this tool</p>	<p>In theory, this tool is completely for developmental purposes, but lacks reliability and validity due to the need to use personality traits and observes behaviour</p>	<p>Being based on personality traits forces this tool to be completely soft criteria based</p>	<p>Very low reliability and validity due to non job specific ratings, as well as it being personality based</p>
<p>Forced-Choice Method 1940s (WWII)</p>	<p>A number of sets of statements, phrases or words describing job performance is presented to a rater. In each set of four statements, two appear favourable, and two unfavourable. However, only one of the favourable statements add to the score, and only one of the unfavourable statements detracts from the score.</p>	<p>Accountability is slightly increased here as it helps to give ratees ideas of proper work functions and behaviours, but the complicated results diminish accountability greatly</p>	<p>This tool is completely developmental, but due to the lack of understanding of the results diminishes the effectiveness of this tool greatly</p>	<p>Due to the nature of this tool, only soft criteria inputs are used</p>	<p>Reliability and validity are very low due to the hard to understand results of the tool</p>
<p>Critical Incident Method 1940s</p>	<p>Originally developed to train pilots in takeoffs and landings. The behaviour, which was crucial to success or failure, were observed and meticulously collected during World War II. In order to use this technique in performance appraisals, supervisors must document positive and negative behavioral events (incidents) that occur during a given performance period.</p>	<p>Accountability is greatly improved with this tool, assuming accountability is built into the instrument</p>	<p>This tool is primarily a judgmental tool used to aid in development. While very effective for judgment due to "observed behaviours", it can lack developmental purposes if not managed properly</p>	<p>Depending on the tools development, it can have a mix of hard and soft criteria. But allows for a possibly strong hybrid approach</p>	<p>Reliability and validity are greatly improved with this tool due to key or "critical" incidents being measured</p>

<p>Management by Objectives (MBO) 1954</p>	<p>This is the first time employees are being appraised based on short-term goals, rather than traits. This process involves multiple steps as listed; 1) The employee is to arrive at a clear statement of responsibilities of his/her position. 2) This statement is reviewed by the manager, and modified until both employee and supervisor agree. 3) The employee then creates a list of goals for the evaluation period. 4) The employee and the manager meet to discuss and modify the goals until both believe they are satisfactory. 5) The final step occurs at the end of the evaluation period or whenever there is a major change in the work situation.</p>	<p>Accountability is extremely high with this tool, as the rater is using goals to determine performance</p>	<p>This tool allows for a significant amount of development based around hard/soft judgment</p>	<p>First tool focus on goals, which will change the amount of hard and soft criteria used, depending on the job, but allows for a strong hybrid approach to be used</p>	<p>Reliability and validity should be extremely high with this tool. However, due to the rather customizable nature of the tool, these may be affected by poor implementation</p>
<p>Behaviourly Anchored Rating Scales (BARS) 1963</p>	<p>Similar to the graphic/trait scales, the BARS approach uses much more accurately defined performance dimensions that are based on observable behaviours. For example, a very high rating for a teacher in the lecture performance dimensions might be “lecturer uses concrete examples to clarify examples”, BARS development is a long and arduous process, involving many steps and many people.</p>	<p>Accountability is in the mid range for this tool, as it strongly depends on the criteria being used. However, due to the need for observed behaviours, accountability will always be judged to some degree.</p>	<p>This tool is primarily used for developmental purposes, with judgmental factors being a byproduct</p>	<p>This tool is designed around using soft input criteria for its results</p>	<p>Increased reliability and validity when executed properly as the measurement dimensions are easy to measure.</p>

9 Appendix B

Table 2: Interview Protocol

Interview Protocol Grounded by Best Practices (with probes indented)		
Interview Questions by Category	Management	Employee
General Strategic		
How would a performance appraisal system benefit the organization, if at all?		
- How will a performance appraisal help with accountability?	*	*
- To Donors/host Government? To the Clients/Beneficiaries? To the sector? To the Organization/staff (quality of judgmental feedback, developmental feedback, perceptions of justice)?		
How would a performance appraisal system hinder this organization?		
- How will a performance appraisal hinder accountability?	*	*
- To Donors/host Government? To the Clients/Beneficiaries? To the sector? To the Organization/staff (quality of judgmental feedback, developmental feedback, perceptions of justice)?		
What barriers have you come across that impeded the implementation of a performance appraisal system?	*	
- How has this organization attempted to overcome these barriers for a performance appraisal system in the past?	*	
Hybrid Criterion Type		
How do you determine that the criteria fully represent the behaviours required to be performed?		
-Why do you take this approach to determining the performance criteria?		
- How do you know that this evaluation was able to help identify your strengths and weaknesses at performing their job?	*	*
How do these criteria help to identify developmental needs?		
Instrumentation (BARS & MBO)		
Please show the kind of form(s) used for performance evaluations.	*	
- Please explain to me why this kind of form is being used.	*	*
- How have employees responded to this form?	*	*

Table 2: Interview Protocol (cont)

Sources of Appraisal		
Who conducts your performance evaluation?	*	
- Did you feel that these appraisers were well informed about the kind of work that you do?		
- Were there other individuals whose participation as raters of your performance would have been helpful? And why?		*
Rater Training		
Were raters qualified for judging your performance?		
- How so?		*
- If you had to make suggestions about how the evaluators could more effectively identify your performance strengths and weaknesses, what suggestions would you make?		
Coaching Style		
Do you think that this evaluation was supportive?	*	*
- Did it help you plan your actions for improvement in the future?	*	*
Bi-Annual Frequency		
How many months ago was your last evaluation?	*	*
How often are evaluations done in this organization?	*	*
- Would you like to see this frequency increased or decreased? Why/Why not?	*	*
Open Ended Questions		
Overall, how comfortable are with having your performance appraisal linked to your pay?	*	*
What would you like to see changed in the evaluation system used at this organization? Why?	*	*

10 Appendix C

Table 3.1A: Accountability (Upward Accountability)

Upward Accountability	
Recognition of Need for Upward Accountability	<ul style="list-style-type: none"> • We are servicing the public, both the donors and the recipients. We are using other peoples money to do things, not our own. We need to change the way we do business, update it. We can't have 4 people taking up half a million in salary. We need to streamline, be more effective in the delivery of our programs. There is too much waste in the system, waste that donors [do not want to] pay for. (M1) • [It] is part of becoming a more professional organization as opposed to a... bleeding heart liberal organization. That has a lot to do with where we are going with performance [appraisals]. Because [the status quo] is not good enough. I mean in the area of education, its books and school and teachers, it's mom and apple pie [and the attitude for the longest time was] how can you possibly not like what we are doing? I think to some degree, this was the culture in which we worked and the standards were just not high enough. It is a whole cultural change that we are trying to bring in... If we want to get better at what we do, and become a more professional organization, we need to have higher standards. If you don't have a good product to offer, then you won't be around very long. We do a really good job, we have a development model, and we can deliver it and implement it better than anybody. And so [with a new performance appraisal system that is] improving that product [(our development model)], it will come to the attention [of the [donors], which will make it easier to raise money. Everything is linked one way or another. Ultimately the strength of the organization is in the strength of its governance. I think the [increased] governance that is taking place has brought a lot of changes, and has improved the quality of the board. When push comes to shove, any donor they want to know, "Did you spend the money?" and the way you show you spent the money is that activities took place. And an audit comes in, they audit you, and they [ask], did these things happen? ... Can you show receipts? So it's more oriented around, that activity level of spending money, and in a timely way. And that you can prove [it]? You can imagine with a funder like CIDA, they are dealing with hundreds of groups. And at the end of the day... they absolutely want to see results, there is no two ways about that. You know, an auditor comes in and looks for receipts, and looks that things were done according to certain deadlines. (M2)
Recognition of Need for Board Accountability	<ul style="list-style-type: none"> • There was a core group of people [on the board] doing most of the work and some individuals who would sit back and not help. So one of the changes, is to bring in board terms [of reference (e.g. operating regulations)], and enforce them, [including having] board transitions. (M2)

Table 3.1B: Accountability (Inward Accountability)

Inward Accountability	
<p>Micro Management/ Time Tyranny</p>	<ul style="list-style-type: none"> • Watching us every minute of the day to make sure we are doing what we are supposed to because it's on [working] hours, that's not right. So if that is what that would do [as a result of a new performance appraisal system], we may as well all quit. It wont be [organization] anymore, it will be something else. (E1) • I have friends in the government who rarely really see their boss, and they can get away with little work, but not here haha. When everyone is so interconnected, it's hard to sit back and do nothing. For the most part that is. (E6) • I believe in an honour system, we work a 7-hour day, and I think that if our people do a hard push, I will let them go a little earlier, [but with the new performance appraisal system] I am not allowed to anymore. Here we are [clocked] by the minute... (M1) • The overtime policy... is a big concern for people. Let's empower our staff to be the best they can be. If you are overseas, you are expected to only work 7.5 hours. Are you supposed to say, oh its 4:30 I have to go home because I am only allowed to work 7.5 hours, give me a break. (M3) • [For the chain of signing authority] You got the director of fundraising, you got the program manager...[then] development [director], [then] the director [of fundraising] again, [then] you have finance director, you have myself, you have the communications officer and ultimately and finally, you have [name of Managing Director] and the grants officer. Too much back and forth. It gets signed off once again once [name of Managing Director] says you finally got it. (M3)
<p>Lack of Procedural Justice</p>	<ul style="list-style-type: none"> • I mean we are all doing the same job essentially, just different programs, why not be evaluated the same. It's like any other company, like working at a store, all cashiers have the same job, their customers served can change dramatically, but it is the same job nonetheless. (E2) • Good management is important. Governance and board is important, HR is very important, [But] for the last three years, for the new corporate policy... the rules are changing while we [already] had an effective one... There [are] always some abusers. We should not penalize the general audience just because of one or two, ... [but] that is what is happening. (M1)

Table 3.1B: Accountability (Inward Accountability) (cont)

Inward Accountability (cont)

Lack of Procedural Justice (cont)

- I think the evaluation process is important, we just had a staff [person who] finished their probation period, and I got the evaluation [done] almost a month and a half after. I breached the agreement by continuous payment [in absence of evaluation] because [the delay in performance appraisal] was not the employees fault. (M1)
- When it was first put in place, there were numerous positions that were really under paid, so they were looking at ways to get people [paid appropriately] without bankrupting the organization, and without creating a lot of jealousies because somebody was getting big pay increases. So... the goal was to show everybody these scales, to show people where they were, to show that everyone was going to be brought up to mid-level. And so... when they first presented it, it seemed quite a positive thing that most people benefited from, significantly. The greatest improvements [due to new performance systems] were with the support staff... [But] with the directors it was different. One, the director of finance who was here for 25 years, there was no way [he would allow] anything... to change, he was near retirement... So again I am not sure if we used our system of evaluation very effectively. [Technically] if you [have] 12 employees and you have 12 forms all signed, then you meet the requirements of the auditor and to some degree its like, ok good we got that looked after... if we are going to truly look at ourselves as professionals, and plan with a level of professional ethos, you can't be making excuses [at any level]. The initial [performance] evaluation [for new hires] focused on the probation period, and in your mind you want certain positions to be taken care of, you don't want to re-open that whole thing. So I probably talked myself into not taking a hard enough analytical look [beyond the probation period] and so that [particular work process] has not improved in any way shape or form. We did set certain clear benchmarks for performance, [but we] really didn't use that for a measure in discussion, as we should have. And I am paying the price for that now, and will have to make some difficult decisions about it. It was interesting how you can allow other events distract you from doing what you need to do in terms of analysis and asking, you know setting up a system and using it without making excuses not to. (M2)

Table 3.1B: Accountability (Inward Accountability) (con't)

Inward Accountability (cont)

Lack of Distributive Justice (cont)

- You know right now it is kind of a mystery as to why we get a pay increase and why we don't. That part really is not clear, it's almost like they just decide randomly to give them out or not. Well that is not true, they gave us all a big one awhile back, but not because we were doing well, it was that we were all being paid so poorly. (E2)
- Really at this point, we only use [the appraisals] to try and move up the pay scale, [but] once you are at the top of your pay scale, there is not a whole lot of benefit towards worrying about them any more to be honest. So maybe [we need] some way of making the evaluations matter more, and be more worthwhile that way. (E5)
- Should there be a salary increase [for new hires] or not. I cannot give it unless the people have been evaluated. [But there] were a few employees screaming that they were being underpaid for their jobs, so a 20% increase was just given to them. This was done without a proper assessment. How is this fair and equitable? How can we do this to 3 people and not everyone else? (M1)
- [Previously] within six months, if you were successful at your job, and you passed your probation, you move to mid-range. But now in the last four or five years, people got up to that [externally] competitive [midrange pay] level, and they are staying with the cost of living, [but] the system is set up so that you really don't move beyond that mid-level. And there was no ability to significantly increase salary unless your job description changed significantly. So if you were doing the same job, and doing it well, you really didn't have much chance to change. [So] pay increases... there hasn't been many of [them] recently. I think, the [new performance appraisal system] system that was put in place, did not lend itself to performance driven pay increases. (M2)
- There is no clarity [about how] to move through the levels. (in relation to promotions) (M3)

Lack of HR Expertise

- Part of the problem here is not having a qualified HR staff member. Do we need a full time staff member? No. But do we need a professional here, yes. If we are supposed to grow... we need a qualified individual to lead the changes to make it happen. This is where we are at right now, we need someone qualified to fix the issues to increase productivity. (M3)

Table 3.2A: General Criterion Issues

Criterion Issues	
General	<ul style="list-style-type: none"> They [CIDA] also sometimes require a work plan that is far too detailed in certain areas and not nearly detailed enough in others. So if we were to solely guide our performance based off of these work plans, you would either have poor performing workers, or a poor performing company. (E5)

Table 3.2B: Criterion Deficiency

Criterion Deficiency	
Staff Numbers Unrecognized	<ul style="list-style-type: none"> You know, [the managing director] is so determined to work off job descriptions, and it's just not right. We do our job descriptions worth and then some... When I started here, we had 3 people in the department and now we have [three times as many] including me. And that is not shown in the job description. It is still the old one, sure we still have the same countries and a few others but I have more than twice the staff. I have the entire organization come knock on my door and ask me to do something, sometimes demand it. And then I can get a lower rating because I am busy doing what I am supposed to in person, but not on paper. How is that fair? [How to] make it work? Make it based off what I do, not what I was supposed to do 8 years ago. I don't know how, but just fix it so it makes sense to my job and can show me what I need to do. (E1)
Mentoring Students Unrecognized	<ul style="list-style-type: none"> The only downside [of the performance appraisal system] is that it is based off our work plans, but as you know, our work plans can change drastically when new opportunities arise... I had an opportunity to have students from Algonquin College come down and help with some of the production of our rollouts. And this was good, but it ate up a lot of extra time that I couldn't focus on other things. So while it was completely worth it, and overall a beneficial action, if it happened around performance appraisal time, it could skew the results cause something may not have gotten done in that past week or two. (E4)

Table 3.2B: Criterion Deficiency (cont)

Criterion Deficiency (cont)	
Qualitative Results Neglected	<ul style="list-style-type: none"> • In theory [using CIDA work plans for our performance objectives] is a great approach. There is [nothing] wrong with what we hand in to CIDA... but the length of them, and level of detail is... not needed. I guess with that said, maybe less emphasis on the CIDA work plans, and more emphasis on actual [individual] performance work plans would also make a big difference [for the utility of the performance appraisals]. No more needing to read between the lines and such. (E5) • Theoretically, [our formal performance appraisals] set targets, and benchmarks and indicators of what constitutes performance.... That is part of the challenge. Just getting your mind around articulating good indicators for example... The challenge for us, [is that] you can have an annual report that says, we distributed 943,832 books, we supported 432 teacher trainer centers, we worked with 1402 libraries... The public, and the public want to know those things, [because] they want to know what you did with their money. But] there are [other important] things you have to do to have a good program; To build partnerships, and to develop meaningful engagement, and to do it in a timely fashion, but [reporting on that] is not always exactly the same as what you have to do to get the money. [So] those are tangible things that they can get their mind around. But [the quantitative metric] doesn't necessarily tell you anything, it doesn't say that, these communities are better off, that these communities have healthier lives, that they are gonna live longer, they are happy, smiling people you know. Its, its um, it's a struggle and I think it has an impact on our ability to, set our own HR targets, and use them in meaningful way. How good a workshop was it? How meaningful were the materials? You know, that doesn't necessarily come through in those types of [quantitative performance appraisal] plans. And that is just the nature of the beast... So I think the challenge for us, is to, subdue [emphasis on quantitative criteria] that, and yet at the same time, develop some [performance appraisal systems and criteria] that are more meaningful to us. I think it can be a lot more useful than we are allowing it to be. I think we are not getting enough out of [CIDA reporting guidelines]... For all good intentions, both sides create these documents, and then end up checking them off, like a checklist, rather than using them for a more meaningful analysis. It's easier to measure quantity and then it is to measure quality. So again, I think the reports are not, as well suited to looking at the qualitative side of any individual's performance. With ten countries, and with different partners, and different projects, you kind of just [listing them off as complete, instead of how they were complete]. [So the CIDA] work plans [were] too much down into the weeds and not enough about wider objectives. (M2)

Table 3.2C: Criterion Contamination

Criterion Contamination	
External Confounds (Partner Problems/Natural Disasters)	<ul style="list-style-type: none"> • We are assessed off the plans we have, [but] the plans we have can change between programs... A partner [(external NGO collaborating on a program)] may drop the ball in one program... So this can show as mediocre performance, where really it comes down to external factors. (E2) • The thing about our business is there are certain things that trump all else. We were about to start up in [country x], but all the best laid work plans, evaluations, procedures couldn't do anything against the fact that the country was destroyed by an earthquake. (M2)

Table 3.3A: Instrumentation

Instrumentation	
Descriptive	<ul style="list-style-type: none"> • I may see a certain process as really important, where he [my manager] may not, or vice versa, and it doesn't mean either one of us are wrong, it is just our perspective. And [the instrument] allows each one to discuss that point of view, and mutually come together to an understanding. (E4) • There is a form. A questionnaire I guess you could classify it as. It's a rather simple process. We have this questionnaire that's a few pages long that asks us our opinion on our performance. (E5) • We have an annual evaluation for all employees. Using their job description as a reference and using the [CIDA] annual work plan as a reference, they [(the employees)] essentially state (according to a certain number of questions) how they feel they performed over the course of that year. Their manager looks at that and adds their own comments... They then meet together to talk about it, and the employee has a chance to write on the document, their take and their feeling about what feedback they got from the manager. (M2)
Rigidity (Incomparable CIDA Goals Across Departments)	<ul style="list-style-type: none"> • One of the things we are looking at is how to have more of a common format for individual work plans [performance objectives] across the entire organization. They require these long very detailed [Result Based Management] output/outcome results statements. The trouble with these documents is they tend to be too [output] activity oriented. So it is page after page of "this workshop will happen, this trip will happen, this report will be written". As a useful document it's too detailed, and too focused on the activity level that doesn't really inform performance. [When translating this into our individual performance appraisal instruments] we need more of a common approach across the organization so that if you were looking at this department, and this department and this department, that they are all apples. You know that it is not different system [being used] in different departments. (M2)
Rigidity (Inflexibility to Amendments)	<ul style="list-style-type: none"> • We [my colleague and myself] came up with the idea of having an annex [to my performance appraisal and job description] to reflect that we have increased to 8 other staff since I started. But my manager later denied this request. (E1)

Table 3.4A: Rater Training (Bias)

Rater Training	
Positive Bias	<ul style="list-style-type: none"> • We have a list of question to answer to basically show how we feel about our own performance. Our manager will then look over it, and then add any ideas or comments to it. We then discuss this, and then we sign off on it. Which is fine...but it really is a focus on what is done well. Rarely do I see anything bad brought up, but maybe that is because we all work well here or maybe it's from [Managing Director] always talking with us so we know where to go...It's hard to say really. But... it is a focus on the positive, which in all reality can only be expected, with a company like [Organization], we want a happy environment, that's why we work for an NGO. I think we focus on the positive, you would have to really miss the ball on your evaluation to get a drastic change in the comments section, or perform really poorly. (E2) • [With] most of the [performance appraisals] I read, it is 95% positive... They are all: "We did this." "This happened." "We did that." It is almost just a listing of great things that happened. There is nothing wrong with that, but rarely do I see any tough aspects come up. Like, "You are really struggling with this element, or you really seem to have a hard time moving ahead on this". I think the thing there is that it is a small organization, and each department is its own group, and I think that every director wants the loyalty of their group, so there is a certain incentive to, not be too critical of anybody or anything. And you can't really criticize anybody outside your group because it is not really your business. So there is a natural, in-born motive to not ask hard question or to not pursue [critical] things too far. (M2)

Table 3.4B: Rater Training (Need for Training)

Rater Training (cont)	
Need for Training	<ul style="list-style-type: none"> • I think [the performance appraisal instrument and process] is a very effective tool for identifying my strengths and weaknesses. But I also think that it can be a very poor tool for it as well. It really comes down to how honest both parties are in the process, and how well trained you are at administering them as well. [In response to the question of what is needed to improve the performance appraisal process] I would say proper training for everyone. Better execution of the system will help both managers and employees fill out these forms better, causing a lot less stress and a lot more functionality. I... see them being easily rushed over, and just kind of checking each one off and getting it done ASAP, but I also see it being done slowly and methodically. One approach will give great results, and other results will give poor results. While [my supervisor] has a little bit of training in the area, he is just a natural leader and coach for the team, and is always there for the extra mile of help. If we didn't have [our manager], [the appraisal discussions] would be far less beneficial. And like I mentioned about other managers previously, I can see their employee's appraisals being rather useless. I don't think there is proper training on how to coach and evaluate and how to correct behaviours. So it comes down to the natural ability of the manager, which in my case makes me lucky, but others may not be. (E5) • The evaluation process here is more of something of a fire stomping process. It is not taken seriously. So I am a little bit rough on these [steps to implement the performance appraisal] with my staff because I [shouldn't] have to remind people, if you are a director you are supposed to assume all of your responsibility. Last year I forced all of the directors to ask the staff to agree [that] the job description is the job they are doing. I think we have all excellent staff, but [organization] needs to improve on this. The personnel policy [for administering performance appraisals] needs to be explained to each director so it can be assimilated equally. They were changed part way through the process back and forth. This was primarily in [relation to] the evaluation process, the pay increases and travel/overtime. And when employees do not know what will happen, it creates uncertainty, which then makes people unhappy. When people are unhappy, they will not be as productive. These people need to know how things are going to happen, and these things need to be consistent. Absolutely, these directors need to be properly trained and coached on how to effectively and consistently, consistently implement policies. This will help make the policies work better. The employees will be happier, the directors will be happier, and the company will be better. I try to be fair with everybody... That way when situations arise, they will know they are being treated fairly. Now there can always be a one time shot to change that, you know in case of an emergency or such, but that should be left up to the director. But yes, all employees need to be treated the same, trained to the same quality, coached to the same quality. This same level of treatment will help create a fair workplace. (M1)

Table 3.5A: Coaching

Coaching	
Supportive Performance Feedback	<ul style="list-style-type: none"> • Well like I just said, it is definite positive situation for the most part, we can get a chance to discuss things that we may have forgotten. Strengths? Yes... I am able to highlight them in the questions given. And then the manager will add... their comments. (E2) • I like that it forces us to sit down and reflect on the work I have done in the past. Talk about what I did well, and what I can do better. Or like, how it shows me where I need to focus on the upcoming projects. (E3) • Once a year... I sit down with my manager, we go over my work plan, and we make notes on it. We talk about where I am doing well, and where I need improvements. For me, it has always been quick and easy. Get it out of the way and get back to work. (E3) • We fill out a standard [self appraisal] form about how we feel we have done, and then we go have that discussion. At which point [Managing Director] will make notes on anything, critique if you will. I think it is a pretty simple process that way. The [supervisor’s feedback on our self-appraisal in the form of] notes really are stopping people from grossly exaggerating what they are doing and such. So yeah we do this, then we talk about how I can improve and look at my work plan and go from there. It’s very much a give and take process, rather than just a verdict being passed down kind of thing. It was nice to sit down and talk about things formally, rather than just the constant... less formal way of a quick pop in meeting. (E4) • [The performance Feedback discussion is] more like where to go in the future with different tasks and such... We discuss the main topics and he will have written down comments upon my evaluation and then we talk about those too. It doesn’t take too long because I work so close with him daily, well probably the closest out of anyone. So really it’s not so much about [judgment of] my [past] performance as much as where to go from here. (E6) • I really feel the evaluation is a chance to tell the staff how well and how [badly] they did their position. What could I have done to assist them better? What would they recommend to the organization to stream line it better? (M1) • I think that it’s an interactive evaluation... that it provides a chance for the employee and manager to sit down and discuss what has been taking place. (M2)
Unhelpful Performance Feedback	<ul style="list-style-type: none"> • I think the main strengths of this process are that it forces a conversation about things that have happened and of things that will [need to] happen. I think the [problematic] issue arises when the quality of that conversation can be drastically different from employee to employee and manager to manager. [My manager] will then go through [our self-appraisals] with us and let know where we could use more detail and obviously [he will express] any disagreement with any statements [that we have made]. But really, [formal performance appraisal process] is more of a formality. (E5) • The non-development [(i.e. the judgmental part of the formal performance appraisal)] that was easy to sort of present and talk about, because the development [part of the discussion] was so far down in the weeds, that, I think that was the problem. I think [the annual performance appraisals don’t] go deep enough on how an employee can be supported to do more and be more effective. And because it is annual, it is less than an informative process it becomes more of a summative process. So it comes down to, “Did you do it or not?” Perhaps a monthly or perhaps a six month [(bi-annual)] process may lend itself to more coaching and looking at what can be brought in, professional training. And [that could more effectively] identify weakness, and [enable the manager to] say you need to go take a course in something, or we all need to take a course in something. (M2)

Table 3.6A: Sources of Appraisal

Sources of Appraisal	
Self Evaluation	<ul style="list-style-type: none"> • Well we have a questionnaire that we will fill out eventually. It asks us about how we did and we write down the way we feel we did. So in that sense it is pretty easy for us to complete. But, you sit there and think, hmmm, what did I do last June? Heck if I can remember... So you fill it in the best you can, and go from there. That will usually get the job done, and anything big will be caught by [Managing Director], or maybe hopefully not haha. • There is a form. A questionnaire I guess you could classify it as. It's a rather simple process. We have this questionnaire that's a few pages long that asks us our opinion on our performance. (E5)
Managerial Evaluation	<ul style="list-style-type: none"> • Then I go to [Manager] and we have a discussion about it, he will help by adding comments and then we sign off on it. (E1) • It's a rather informal process. I sit down with who I report to, which happens to be [Managing Director]. We have a discussion based on my performance. (E4)
Self/Managerial Evaluation	<ul style="list-style-type: none"> • Once a year I will fill out a self-evaluation form. And then I hand it into [Managing Director] where he will review it, and then we meet about it. (E6)

Table 3.7A: Frequency (Onerous Overlap)

Frequency	
Descriptive	<ul style="list-style-type: none"> Yearly, every March or April or May. But yes, they [formal performance appraisals] happen only once a year. Which... is also a positive. Having to deal with them [only] once a year is nice because we can get so busy. But at the same time it can take some effort to think back to how things were done 9-10-11 months ago. That's why I find good note taking [on an ongoing basis] can make the difference. I find it helps. (E1) [It happens] once a year from my understanding. I had mine just a few months ago, it was good. (E4)
Onerous Overlap	<ul style="list-style-type: none"> The [annual performance appraisals] are usually done right around the reporting [to donors] time of the year, so the last thing we want to do is another report by ourselves commenting on ourselves... (E1) [The timing of the formal performance appraisals] is not always the most convenient of times. So maybe [it would be better to] do it when we have a slower period... But does it always have to be 12 months exactly? If not, let it slide by a month or two to make things better. (E3) [The appraisals take place] every year. Usually in the spring or so. Just depends on how busy reporting season is you know haha. No one wants to do evaluations when there are government deadlines or a huge mail out being prepared. (E3) [Last year it took place in the] spring, probably later than it was supposed to be. But the work is so ebb and flow, that at times it gets way too busy with normal job functions that we don't have time to properly do the formal process of the evaluation. (E5) To really make use of [the performance appraisal feedback process] though, we've got to have more of a day to day, if not... [at least a] more regular approach [to] performance management, as opposed to once a year. We are still looking at it at the end of the year. And by that time, of course it is too late. Our annual budget process is in February, and [we need] to put forward a budget... at the March Board meeting to have it approved by the beginning of the new fiscal year, April 1st. [So] if budget and employee compensation was linked to job performance, then you would want to have those evaluations done, say in December or January, so that would factor into the budget process. And instead we tend to do it, we tend to get everything else out of the way, and in April we finally get around to it. (M2)

Table 3.7B: Frequency (Formal Coaching Frequency)

Frequency (cont)	
<p>Coaching Frequency Compliments Formal Frequency</p>	<ul style="list-style-type: none"> • There is an annual evaluation, so once a year we will be evaluated. Now with that said, I do find a lot of help comes from [Director]. You know a constant discussion of how things are going, and where things are going. His door is always open, even when its not, you can knock. So we are always getting feedback and going back and forth with each other so we can offer the best programs possible. (E2) • Either you do your job or you don't. Either you answer the phones or you don't. And [Managing Director] was always good about giving me tips as we go. So I always found [doing my performance appraisal to] be an easy thing. (E3) • As a somewhat new individual to this role, [my supervisor's] constant coaching... helps me understand what is needed from me, and what I am doing well. This is what really helps. Sure the yearly evaluation is helpful, but in all honesty, it's the water cooler chats and the quick comments about something you did well or not that works. He [my manager] would rather make small adjustments constantly to the direction of the "ship" than one or two large corrections a year where you have allowed yourself to go way off course. And let's be honest, this is a far better approach to use that a simple self-appraisal sheet based on some sometimes-arbitrary goals. (E5) • Formally they [the performance appraisals] are done annually, but in actuality it is almost a constant evaluation or conversation about performance... If you are told everything is going well... [then] sudden [concerns about performance] should be gradual. (E5) • I guess my job is a little different because of the fact that I am an executive assistant. I prepare letters, answer some correspondence. So when it comes to the performance aspect I usually get feedback daily, more than daily. So I can't do a week's worth of work without being critiqued, I can barely go half a day some days. It's not that he is always telling me I am doing something wrong. It is more just how to do it more so [that] he would like it. So it is good, I guess it is like a constant level of feedback. (E6) • They are quick and easy. I guess they don't really get in the way too much, which is nice. You know when an organization is this small, I don't think you need a huge drawn out process because you are getting feedback all the time. (E6) • [Performance appraisals are done] yearly. Usually in the spring. But it depends, it can vary, it's not like it's every May 23rd. Its more like every year after the reporting and annual work plan is done for CIDA, and then when [the MD] isn't away. [So] yes and no [to the utility of the performance appraisal for me]. Like I said, because I work closely with [management] daily, I kind of get that [performance] feedback daily. But it [(formal performance appraisal)] does help sometimes to pinpoint issues with the work plan that is set up and how we execute. (E6)